

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 28, 2004, 12:44:51 ; Search time 820 Seconds
(without alignments)
10672.302 Million cell updates/sec

Title: US-09-831-804-1

Perfect score: 2060
Sequence: 1 cttattagggaagtgtgct.....gacatactctcttaatg 2060

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: N_genseq29Jan04:*
2: genseq1990s:*
3: genseq2000s:*
4: genseq2001as:*
5: genseq2001bs:*
6: genseq2002s:*
7: genseq2003as:*
8: genseq2003bs:*
9: genseq2003cs:*
10: genseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|---------------------|
| 1 | 2060 | 100.0 | 2060 | 3 | AA15398 DNA encod |
| 2 | 1235.8 | 60.0 | 1239 | 6 | AB232206 Candida a |
| 3 | 99.4 | 4.8 | 2132 | 7 | ADA53158 Human cod |
| 4 | 98.4 | 4.7 | 1705 | 8 | ACA98970 Human enco |
| 5 | 96.6 | 4.7 | 2320 | 7 | ACA56456 Human sig |
| 6 | 95.2 | 4.6 | 4563 | 5 | AA591317 DNA encod |
| 7 | 94.6 | 4.6 | 2597 | 7 | AA151569 Human nuc |
| 8 | 92.4 | 4.5 | 3639 | 5 | AA664586 DNA encod |
| 9 | 91.8 | 4.5 | 2110 | 9 | ADB63624 Human cod |
| 10 | 91.8 | 4.5 | 2760 | 9 | ADC56695 Human mac |
| 11 | 91.8 | 4.5 | 3078 | 9 | ADB62468 Human cod |
| 12 | 90.4 | 4.4 | 2026 | 7 | AA55863 Human nuc |
| 13 | 90.4 | 4.4 | 2114 | 7 | ADA53124 Human cod |
| 14 | 90.2 | 4.4 | 2230 | 7 | ADA52931 Human cod |
| 15 | 90 | 4.4 | 1952 | 9 | ADC58104 Zinc fing |
| 16 | 89.2 | 4.3 | 2064 | 9 | ADC30762 Human nov |
| 17 | 89.2 | 4.3 | 2622 | 5 | AA568872 DNA encod |
| 18 | 89.2 | 4.3 | 2723 | 6 | AA16178 Human cod |
| 19 | 89.2 | 4.3 | 3839 | 6 | AAK83826 Human cod |
| 20 | 89 | 4.3 | 2905 | 5 | AAH16608 Human CDN |
| 21 | 89 | 4.3 | 2905 | 5 | AA566143 DNA encod |
| 22 | 89 | 4.3 | 3020 | 5 | AA592560 DNA encod |
| 23 | 89 | 4.3 | 3502 | 4 | AA157845 Human pol |

| | | | | | |
|----|------|-----|------|---|--------------------|
| 24 | 88.6 | 4.3 | 2597 | 7 | AA55855 Human nuc |
| 25 | 88.6 | 4.3 | 4227 | 8 | ACA98938 cDNA enc |
| 26 | 88.4 | 4.3 | 2509 | 6 | AA31103 Human tra |
| 27 | 87.6 | 4.3 | 1757 | 6 | ABQ81135 TRAF6-Inh |
| 28 | 87.2 | 4.2 | 831 | 6 | ABQ5095 Human ova |
| 29 | 87 | 4.2 | 976 | 5 | AA569188 DNA encod |
| 30 | 86.8 | 4.2 | 2662 | 6 | ABQ93353 Human cod |
| 31 | 86.8 | 4.2 | 6219 | 7 | ACC46324 Human dit |
| 32 | 86.8 | 4.2 | 6316 | 7 | ACC46373 Human dit |
| 33 | 86.2 | 4.2 | 1777 | 7 | ABX34443 Human mdd |
| 34 | 86.2 | 4.2 | 2558 | 4 | AA526690 Human gen |
| 35 | 86.2 | 4.2 | 2558 | 4 | AA526691 Human gen |
| 36 | 86.2 | 4.2 | 2558 | 7 | ABX74039 Human nov |
| 37 | 86.2 | 4.2 | 2558 | 7 | ABX74040 Human nov |
| 38 | 86.2 | 4.2 | 2607 | 5 | AA587125 DNA encod |
| 39 | 86.2 | 4.2 | 2681 | 7 | ABX34772 Human mdd |
| 40 | 86 | 4.2 | 1890 | 9 | ADC30336 Human nov |
| 41 | 86 | 4.2 | 3309 | 7 | ACC46347 Human dit |
| 42 | 86 | 4.2 | 3400 | 9 | ADB62883 Human cod |
| 43 | 85.6 | 4.2 | 1549 | 5 | AA567562 DNA encod |
| 44 | 85.6 | 4.2 | 2476 | 7 | ACC46407 Human dit |
| 45 | 84.8 | 4.1 | 2239 | 7 | ADA53516 Human cod |

ALIGNMENTS

RESULT 1

AA15398
ID AA15398 standard; DNA; 2060 BP.

AC AA15398;

DT 04-SEP-2000 (first entry)

DE DNA encoding a transcription factor designated CATP11A.

KW Transcription factor; CATP11A; DNA-binding protein;

KW ribosomal RNA 5S gene; fungal infection; ss.

XX Candida albicans.

XX Key Location/Qualifiers

FT CDS 720..1958

FT /tag= a

FT /transl_except= (pos: 1296..1298, aa: Ser)

FT /transl_except= (pos: 1734..1736, aa: Ser)

PN WO200028037-A1.

PD 18-MAY-2000.

PF 09-NOV-1999; 99WO-FR002739.

PR 10-NOV-1998; 98FR-00014147.

PA (HMRI) HOECHST MARION ROUSSEL.

PI Bordon-Pallier F, Camier S, Sentenac A;

DR WPI; 2000-376549/32.

DR P-PSDB; AAY93316.

PT New nucleic acid encoding Candida albicans transcription factor, useful e.g. in screening for antimycotic agents and for immunization.

XX Claim 4; Page 32-33; 45pp; French.

XX The present sequence encodes a Candida albicans transcription factor, designated CATP11A. The polypeptide is a DNA-binding protein, which is involved in initiating transcription of the ribosomal RNA 5S gene. The polynucleotide is used to screen for its specific inhibitors, potentially useful as antimycotic agents, to raise an antibody response that is

CC protective against fungal infection and to raise antibodies. Such
CC antibodies, as well as the polypeptides and polynucleotides are used in
CC compositions for diagnosing and treating fungal infections, e.g. by
CC detecting polymorphisms and mutations

XX Sequence 2060 BP; 726 A; 315 C; 354 G; 665 T; 0 U; 0 Other;

Query Match 100.0%; Score 2060; DB 3; Length 2060;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2060; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 CTTTATAGAGATGGCTGAGGCGATTTGTATTCACGGGCTCCCAAGTGAATGTTT 60
DB 1 CTTTATAGAGATGGCTGAGGCGATTTGTATTCACGGGCTCCCAAGTGAATGTTT 60
QY 61 TAGTAATATCCATCATTTGGGCTCAGTGTGAATGGGGTGTGTCATCTCTTGTGTAG 120
DB 61 TAGTAATATCCATCATTTGGGCTCAGTGTGAATGGGGTGTGTCATCTCTTGTGTAG 120
QY 121 AAATAGGCGAGGCTCCGATCCCAAAAAAGAAAGATCAGATGTCCTGGCTGCAAGA 180
DB 121 AAATAGGCGAGGCTCCGATCCCAAAAAAGAAAGATCAGATGTCCTGGCTGCAAGA 180
QY 181 TTTGTGACCATGGCAATGCCGAAAAATGAAAAAAGTCTACTGGGCCCACT 240
DB 181 TTTGTGACCATGGCAATGCCGAAAAATGAAAAAAGTCTACTGGGCCCACT 240
QY 241 ACAAAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
DB 241 ACAAAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
QY 301 TATTTGACGGGCTTTAGAAATTTGTATATTTGTGTCTGACACTCTGTGTTAATCT 360
DB 301 TATTTGACGGGCTTTAGAAATTTGTATATTTGTGTCTGACACTCTGTGTTAATCT 360
QY 361 GGAACATGCTGCTCCCTGTGAAAGGCTGCTGTAATGAATCATGATCAAGATATAT 420
DB 361 GGAACATGCTGCTCCCTGTGAAAGGCTGCTGTAATGAATCATGATCAAGATATAT 420
QY 421 GACTTGCTCACTTACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
DB 421 GACTTGCTCACTTACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
QY 481 TATGTAACCACTTGACTTATTTCTTGTGTGGATTCATTTGGATGATGATGATGAT 540
DB 481 TATGTAACCACTTGACTTATTTCTTGTGTGGATTCATTTGGATGATGATGATGAT 540
QY 541 CAATGTAATTTGGTACCAATCTTTTGTGCTGTGCGGACTTCCTTAATATGCGACT 600
DB 541 CAATGTAATTTGGTACCAATCTTTTGTGCTGTGCGGACTTCCTTAATATGCGACT 600
QY 601 ATTTGATTAATGAGACGCAAGCATTCCTCTTCATAGAAAAAATPAACAACCTGA 660
DB 601 ATTTGATTAATGAGACGCAAGCATTCCTCTTCATAGAAAAAATPAACAACCTGA 660
QY 661 AAAAATTAACAGCGGACTCATCTCTTTTCAAACTTTTAACTTTTAACTTTTAACT 720
DB 661 AAAAATTAACAGCGGACTCATCTCTTTTCAAACTTTTAACTTTTAACTTTTAACT 720
QY 721 TGAATGAAAGTGAAGAACCAATCGATATCATCTTTAATATCTCTCTTCACTAC 780
DB 721 TGAATGAAAGTGAAGAACCAATCGATATCATCTTTAATATCTCTCTTCACTAC 780
QY 781 GTCCCAAAAAGTATATTTGACATATGAAAGGTGTGATAAGCCTATATGACCATCAT 840
DB 781 GTCCCAAAAAGTATATTTGACATATGAAAGGTGTGATAAGCCTATATGACCATCAT 840
QY 841 TATTAAGCAACATTTAAGAACCAAGATATGATGACCGGTATTAATGATGATGACG 900
DB 841 TATTAAGCAACATTTAAGAACCAAGATATGATGACCGGTATTAATGATGATGACG 900
QY 901 ATTGTGTAAGCAATTTTGAAGAAATCATTTGAAACCATATGATGATGATGATGATG 960
DB 901 ATTGTGTAAGCAATTTTGAAGAAATCATTTGAAACCATATGATGATGATGATGATG 960
```

```
DB 901 ATTGTGTAAGCAATTTTGAAGAAATCATTTGAAACCATATGATGATGATGATGATG 960
QY 961 AAAAAAACCATTCATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
DB 961 AAAAAAACCATTCATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
QY 1021 AAAGCATGAATTCACCCGATACAAAGTCAATTAATGATGATGATGATGATGATGATG 1080
DB 1021 AAAGCATGAATTCACCCGATACAAAGTCAATTAATGATGATGATGATGATGATGATG 1080
QY 1081 CATTTATTAACATCATCTTTAAGCATCATATATATCTGTGATGAAAAAATGATTA 1140
DB 1081 CATTTATTAACATCATCTTTAAGCATCATATATATCTGTGATGAAAAAATGATTA 1140
QY 1141 CGGTAAACAAATGATTAAGTTTTCACGACCTTCAAAATGAGCAACATTAATTA 1200
DB 1141 CGGTAAACAAATGATTAAGTTTTCACGACCTTCAAAATGAGCAACATTAATTA 1200
QY 1201 AACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1260
DB 1201 AACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1260
QY 1261 AAACCTGTCAGTATTTACAAATTTGATTAACAACTGATTCACAAATGATGATGAT 1320
DB 1261 AAACCTGTCAGTATTTACAAATTTGATTAACAACTGATTCACAAATGATGATGAT 1320
QY 1321 AATGTGTAAGGTTGTGTGGAAAAAGTTTACCTCAATGATGATGATGATGATGATGAT 1380
DB 1321 AATGTGTAAGGTTGTGTGGAAAAAGTTTACCTCAATGATGATGATGATGATGATGAT 1380
QY 1381 ATTCTACATGATCAAAATATGATGATGATGATGATGATGATGATGATGATGATGAT 1440
DB 1381 ATTCTACATGATCAAAATATGATGATGATGATGATGATGATGATGATGATGATGAT 1440
QY 1441 AAAATGAATTTGTTGAACATTAATATATCTTCATGATGATGATGATGATGATGAT 1500
DB 1441 AAAATGAATTTGTTGAACATTAATATATCTTCATGATGATGATGATGATGATGAT 1500
QY 1501 TAAAGAACTGAAGTGAAGAAAAATTAAGAACTATTAAGTCAAGATGATGATGAT 1560
DB 1501 TAAAGAACTGAAGTGAAGAAAAATTAAGAACTATTAAGTCAAGATGATGATGAT 1560
QY 1561 ATTGCAATGATTAAGAAACAGAGAAATTAAGTGAAGAAATGAAAGATGAAAG 1620
DB 1561 ATTGCAATGATTAAGAAACAGAGAAATTAAGTGAAGAAATGAAAGATGAAAG 1620
QY 1621 ATAGCTAGATGAAAAAGAGTGAATGATGATGATGATGATGATGATGATGATGAT 1680
DB 1621 ATAGCTAGATGAAAAAGAGTGAATGATGATGATGATGATGATGATGATGATGAT 1680
QY 1681 TAAATCATTTACTGCTCTTGAAGGTTCAAGAGTGTTCCTAACTTATCTGATA 1740
DB 1681 TAAATCATTTACTGCTCTTGAAGGTTCAAGAGTGTTCCTAACTTATCTGATA 1740
QY 1741 GTGGAAAGAGATCAATTTGCTAAGAAATATGATGATGATGATGATGATGATGATG 1800
DB 1741 GTGGAAAGAGATCAATTTGCTAAGAAATATGATGATGATGATGATGATGATGATG 1800
QY 1801 ATTTAGCTGACATTTGAATGCGATGATGATGATGATGATGATGATGATGATGAT 1860
DB 1801 ATTTAGCTGACATTTGAATGCGATGATGATGATGATGATGATGATGATGATGAT 1860
QY 1861 ATAGTATGAAAAAGAGAACTCCAGAGGTGAACCATGTTGTAAGAAAGCCAGATG 1920
DB 1861 ATAGTATGAAAAAGAGAACTCCAGAGGTGAACCATGTTGTAAGAAAGCCAGATG 1920
QY 1921 ATTATTTGCAAAATGAACATCAGTATTTCTGATATATATCAATTTAATATTA 1980
DB 1921 ATTATTTGCAAAATGAACATCAGTATTTCTGATATATATCAATTTAATATTA 1980
QY 1981 CATTTTATTTCTTTAATTTTATTTTGTGGCTTTTATTTTATTTTATTTTATTT 2040
DB 1981 CATTTTATTTCTTTAATTTTATTTTGTGGCTTTTATTTTATTTTATTTTATTT 2040
```

QY 2041 GACATATTTACTCTTTAATG 2060
 Db 2041 GACATATTTACTCTTTAATG 2060

RESULT 2
 ABZ32206
 ID ABZ32206 standard; DNA; 1239 BP.

AC ABZ32206;
 DT 30-JAN-2003 (first entry)
 DE Candida albicans essential gene SEQ ID NO 6493.

XX Fungus; Yeast; tetracyclin; promoter; GRACE strain; biosynthesis;
 KM signal transduction; DNA replication; cell division; growth;
 XX proliferation; Candida albicans; fungicide; antifungal; gene; ss.

OS Candida albicans.

PN WO200253728-A2.

PD 11-JUL-2002.

PF 26-DEC-2001; 2001WO-US049486.

PR 29-DEC-2000; 2000US-0259128P.

PR 20-FEB-2001; 2001US-00792024.

PR 22-AUG-2001; 2001US-0314050P.

PA (ELIT-) ELITRA PHARM INC.

PI Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;

DR MPI: 2002-S66694/60.

XX P-PSDB; ABP73656.

PT Constructing strains for identifying gene products as effective targets
 for therapeutic intervention, by inactivating in the strain one allele of
 a gene and placing other allele of the gene under conditional expression.

PS Claim 37; SEQ ID NO 6493; 167bp + Sequence Listing; English.

XX The invention relates to constructing (M1) a strain of diploid fungal
 cells in which both alleles of a gene are modified, comprising modifying
 one allele by insertion or replacement by a cassette having an
 expressible selectable marker and modifying other allele by
 recombination, of a promoter replacement fragment with a heterologous
 promoter, so that expression of the second allele is regulated by the
 promoter. (M1) is useful for constructing a strain of diploid fungal
 cells in which both alleles of a gene are modified. The diploid fungal
 cells having both alleles modified are useful for identifying a gene that
 is essential to the survival or growth of a fungus, a gene that
 contributes to the virulence and/or pathogenicity of a fungus, a gene
 that contributes to the resistance of a diploid fungus to an antifungal
 agent, an antifungal agent that inhibits the growth of a diploid fungus
 and for identifying a therapeutic agent for treatment of a mammalian
 disease. (M1) is useful for identifying a compound which modulates the
 activity of a gene product, preferably enzymatic activity, carbon
 compound catabolism, biosynthetic, transporter, transcriptional,
 translational, signal transduction, DNA replication and cell division
 activity. The method is useful for identifying a compound having the
 ability to inhibit growth or proliferation of C. albicans cells and for
 treating infection by C. albicans. The present sequence is that of an
 essential Candida albicans gene used in the method of the invention.
 Note: The sequence data for this patent is not represented in the printed
 specification but is based on sequence information supplied to Derwent by
 the European Patent Office

Sequence 1239 BP; 481 A; 171 C; 215 G; 372 T; 0 U; 0 Other;

Query Match 60.0%; Score 1235.8; DB 6; Length 1239;
 Best Local Similarity 99.8%; Pred. No. 9.7e-217;
 Matches 1237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

| | | | |
|----|------|--|------|
| QY | 720 | ATGAGTGAAGTGCAGAAACCAATGATATCATCTTATATATCTTCTTCATCA | 779 |
| Db | 1 | ATGAGTGAAGTGCAGAAACCAATGATATCATCTTATATATCTTCTTCATCA | 60 |
| QY | 780 | CGTCCCAAAAGTATTTTGCACATATGAAAGGTGATTAAGCTTAATGCAACATCA | 839 |
| Db | 61 | CGTCCCAAAAGTATTTTGCACATATGAAAGGTGATTAAGCTTAATGCAACATCA | 120 |
| QY | 840 | TTATTGAGCAACATTTAAGAACCCAGTAATGATCGACCGTATTAAGTAAAGTGAC | 899 |
| Db | 121 | TTATTGAGCAACATTTAAGAACCCAGTAATGATCGACCGTATTAAGTAAAGTGAC | 180 |
| QY | 900 | GATTTGATTAAGCATTTTTCAGAAATGCAATTTGGAACATATTTGATCATTCC | 959 |
| Db | 181 | GATTTGATTAAGCATTTTTCAGAAATGCAATTTGGAACATATTTGATCATTCC | 240 |
| QY | 960 | GAAGAAAAACCATTCATTTGTCAGTGTGTGTAAGGGGTTAATTTCTGACAACTTG | 1019 |
| Db | 241 | GAAGAAAAACCATTCATTTGTCAGTGTGTGTAAGGGGTTAATTTCTGACAACTTG | 300 |
| QY | 1020 | AAAAGCATGAATCACCATACAAAGTCAATTTAATGTACATTTGAAAATTTGTCAAGA | 1079 |
| Db | 301 | AAAAGCATGAATCACCATACAAAGTCAATTTAATGTACATTTGAAAATTTGTCAAGA | 360 |
| QY | 1080 | GCATTTTATTAACATCAATCTTTAAGACATCATATATATCTGTCATGAAAAACATTA | 1139 |
| Db | 361 | GCATTTTATTAACATCAATCTTTAAGACATCATATATATCTGTCATGAAAAACATTA | 420 |
| QY | 1140 | ACGTGTAAACATGTATTAATTAATTTTCACTGCACCTTCAAAATTTGCACATTAATTA | 1199 |
| Db | 421 | ACGTGTAAACATGTATTAATTAATTTTCACTGCACCTTCAAAATTTGCACATTAATTA | 480 |
| QY | 1200 | AAACATCATGTGATCTCCTCTTATCAATGATGATCATCCGGTTTAAAAAATTTTC | 1259 |
| Db | 481 | AAACATCATGTGATCTCCTCTTATCAATGATGATCATCCGGTTTAAAAAATTTTC | 540 |
| QY | 1260 | CAACTTGTGATGATTTAATCAATTTAATTAACAAACATGATCAAAACCTTAATGCTCT | 1319 |
| Db | 541 | CAACTTGTGATGATTTAATCAATTTAATTAACAAACATGATCAAAACCTTAATGCTCT | 600 |
| QY | 1320 | AAATGTGTAAAGGTTGTGTTGGGAAAAAGTTTATCTTCAATGTGTAATGATCAAT | 1379 |
| Db | 601 | AAATGTGTAAAGGTTGTGTTGGGAAAAAGTTTATCTTCAATGTGTAATGATCAAT | 660 |
| QY | 1380 | GATTTACCATGATCAAAATATGACCTGTGATTAATTTGATGATGGGGAAATTTGCAAG | 1439 |
| Db | 661 | GATTTACCATGATCAAAATATGACCTGTGATTAATTTGATGATGGGGAAATTTGCAAG | 720 |
| QY | 1440 | AAAAATGAATTAAGTTGAACATTATTAATATCTTCATGATGATGATATCCCTGATATTA | 1499 |
| Db | 721 | AAAAATGAATTAAGTTGAACATTATTAATATCTTCATGATGATGATATCCCTGATATTA | 780 |
| QY | 1500 | TTAAAGGAACGTAAGTAAAAATTTAGAGAACCTTATTAATGATCAAGATTCGAATTAAT | 1559 |
| Db | 781 | TTAAAGGAACGTAAGTAAAAATTTAGAGAACCTTATTAATGATCAAGATTCGAATTAAT | 840 |
| QY | 1560 | AATTTGATGATTAAGAAACAGAGAAATTTAAAGTGAAGAAAGATGAAGAGATGAAGA | 1619 |
| Db | 841 | AATTTGATGATTAAGAAACAGAGAAATTTAAAGTGAAGAAAGATGAAGAGATGAAGA | 900 |
| QY | 1620 | GATGTCATGATGAAAAAGAGTGAATTAAGATGATGATGATGATGATGATGATGATGAT | 1679 |
| Db | 901 | GATGTCATGATGAAAAAGAGTGAATTAAGATGATGATGATGATGATGATGATGATGAT | 960 |
| QY | 1680 | ATTAATATCATTTACCTGCTTTTGAAGAGTTCAAGAGTGTCTTAAACCTTATTTGAT | 1739 |
| Db | 961 | ATTAATATCATTTACCTGCTTTTGAAGAGTTCAAGAGTGTCTTAAACCTTATTTGAT | 1020 |
| QY | 1740 | AGTGGGAAAGAGATCAATTTGTCTTAAGATTAATTTGTATAGAAATGTTTCTAGAGATAT | 1799 |

```

DB      1021 AGTGGAAAGAAATCATATGTCCTAAGAAATATGATAGAAATGTTTCTAGAGATAT 1080
QY      1800 GATTACGTCGACATTTGAATGATGATGATTAATTTCAAGAATTTAGTCATTTCTTA 1859
DB      1081 GATTTACGTCGACATTTGAATGATGATGATTAATTTCAAGAATTTAGTCATTTCTTA 1140
QY      1860 AATAGTATGAAAAAGAAAGAACTCCAGAAAGTGAACCATTTGGTTAAAAAAGCCAGGATG 1919
DB      1141 AATAGTATGAAAAAGAAAGAACTCCAGAAAGTGAACCATTTGGTTAAAAAAGCCAGGATG 1200
QY      1920 GATTATTTGCCAAATGAAAACATCAGTGAATTTCTCGATAA 1958
DB      1201 GATTATTTGCCAAATGAAAACATCAGTGAATTTCTCGATAA 1239

```

RESULT 3

ADAS3158 standard; cDNA; 2132 BP.

ADAS3158;

20-NOV-2003 (first entry)

Human coding sequence, SEQ ID 726.

Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
Gene Therapy; human; secretory protein; membrane proteins; cancer;
inflammatory disease; osteoporosis; neurological disease; gene; ss.

Homo sapiens.

EP1293569-A2.

19-MAR-2003.

21-MAR-2002; 2002EP-00006586.

14-SEP-2001; 2001JP-00328381.

24-JAN-2002; 2002US-0350435P.

(HELI-) HELIX RES INST.

(REAS-) RES ASSOC BIOTECHNOLOGY.

Iisogai T, Sugiyama T, Otsuki T, Makamatsu A, Sato H, Ishii S;
Yamanoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y;

WPI: 2003-395539/38.

P-PSDB; ADA54797.

New polynucleotides encoding full-length polypeptides, e.g. secretory
and/or membrane proteins, useful for developing medicines for diseases in
PT which the gene is involved, or as target molecules for gene therapy.

PS Claim 1; SEQ ID NO 726; 205pp; English.

The present invention relates to novel human secretory or membrane
proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
ADA54071). The coding sequences are useful in the gene therapy of
CC diseases caused by abnormalities of the proteins, e.g. cancer,
CC inflammatory diseases, osteoporosis or neurological disease.

Sequence 2132 BP; 833 A; 423 C; 347 G; 529 T; 0 U; 0 Other;

Query Match 4.8%; Score 99.4; DB 7; Length 2132;

Best Local Similarity 50.6%; Pred. No. 6.6e-09;

Matches 297; Conservative 0; Mismatches 281; Indels 9; Gaps 2;

QY 726 GAAAGTACGAAACCAATGATATCATCTTTATATCTTTCTTCTGATCAGTCC 785
DB 1384 GAATGTGGCAAAAGCTTTTAAACGAGTCCATCTTACTACTACATAAGAGATTCTACT 1443

```

QY      786 AAAAGTATATTTGCACATATGAAAGGTTGATAAAGCCCTATATCGACCATATTATTA 845
DB      1444 GGAGAGAAATTCCTACAAATGTGAAGATGTGGCAAGCTTCTATCGATCCCAAACTT 1503
QY      846 GAGCAATTTTAAGAACCCACAGTATGATGACCGGTATTAATGTAACAGTGAAGATTGT 905
DB      1504 ACTGAACATPAGAAAATTCATCTCTGAGAGAAACCCTACACAGT-----GAGAGATTGT 1557
QY      906 GATTAAGCATTTTTCAGAAAATCACTTTGGAACGATATGATATCATTCGCAAAAA 965
DB      1558 GGCAAAAGCCTTTTAAACCACTCTCAACCTTGCTACACATPAGATTAATTCATCTCGAGAG 1617
QY      966 AAACATTCATTTGTTCAGTGTGTGTAAAGGGTTTAATTTTCGACACACCTTGAAGA 1025
DB      1618 AAACCTTACCAATGTGAAGAAATGTGTAAAGCCTTTAAACAGTCTCAACCTTACTAGA 1677
QY      1026 CATGAATCAACCCATTCGAAAGTCAATTAATGTACATTTGAAAATTTGTCAAGAGCATTT 1085
DB      1678 CATAGAGAAATTCATPACTGTGAGAGAAACCCTACCAATGTGAAAAAATGTGCAAAAGCTTTT 1737
QY      1086 TATTAACATCAATCTTTAA--GACATCAATATATATCTGTTCATGAAAAAAGATTAAAG 1142
DB      1738 AACAGTCTCTCAAACTTACTGTGACATTAAGAAAATTCATCTGTGAGAAACTCTACAAA 1797
QY      1143 TGTAAACAATGTATATAAGTTTCACTGACCTTCAAAATTTAGCACACATAAATTBAAA 1202
DB      1798 CTTAAAGATGTATACAGTGTATTTGAAAACACTTCAAAAGTTTCTAAACATTAAGAAAT 1857
QY      1203 CATCATGTGTGATCTCTCTGTTTCAATGTGATCAATCTGTGTTTAAATTTCCAA 1262
DB      1858 TATGCTGTGTAGAAATCTTAAAGAAATGTGAAGAAATGTGAACAAACCTTTAAAGTTGTAC 1917
QY      1263 ACTTGTGAGTATTTACATTTTCATATAAAACAATGCATCCAAAGT 1309
DB      1918 ACTTGAATTTGCTAATAGATTAATTCATCTPAAAAAACCCTAATAGT 1964

```

RESULT 4

ACA98970 standard; cDNA; 1705 BP.

ACA98970;

25-JUL-2003 (first entry)

cDNA encoding human nucleic acid-associated protein (NAAP) #51.

Human; nucleic acid-associated protein; cytosolic; antitumor; anti-
anticovulsant; nootropic; neuroprotective; cerebroprotective; anti-HIV;
antiallergic; antiinflammatory; chymotrypsin; gene therapy;
cell proliferative disorder; cancer; atherosclerosis;
neurological disorder; epilepsy; Huntington's disease; stroke;
immune disorder; inflammatory disorder; AIDS; allergy;
developmental disorder; Hypothyroidism; Cushing's syndrome; infection;
protein-protein interaction; drug-target interaction;

Homo sapiens.

WO2003023003-A2.

20-MAR-2003.

05-SEP-2002; 2002WO-US028540.

07-SEP-2001; 2001US-0317792P.

07-SEP-2001; 2001US-0317912P.

14-SEP-2001; 2001US-0322270P.

21-SEP-2001; 2001US-0324040P.

28-SEP-2001; 2001US-0326732P.

19-OCT-2001; 2001US-0346716P.

25-JAN-2002; 2002US-0351749P.

22-FEB-2002; 2002US-0359498P.

QY 788 AAGTATATTGACATATGAGGCTGTGATTAAGCTATATGACCATATTATTAGA 847
DB 964 AGAGAAACCTTACAAATGTAAAGATGTGTAAAGCTTTTAAACGATCTTCAACCTTTAC 1023
QY 848 GCAACATTTAAGAACCCAGATATGATGACCGTATTAATGTACAGTGGACATTTGTA 907
DB 1024 TACCATAGAAAAATTCATCTGAGAGAAAACCTTACAAATGT-----GAAAGATGTG 1077
QY 908 TAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTTGATTCATTCGAAAAAAA 967
DB 1078 CAAGCCTTTAAGAGATCCTCAAACTTACTACACATTAAGATTAATTCATCTGAGAGAA 1137
QY 968 ACCATTCATTTGTCAGTGTGTGTAAAGGGTTAATTCGACACACCTTGAAGAGCA 1027
DB 1138 ACCCTCAAAATGTAAAAATGTGAAAAAGCCTTTAACAGCTGTGCACACCTTACACACA 1197
QY 1028 TGAATCACCCTATACAAAGTCATTTTAAATGTACATTTGAAATTTGTCAAGACATTTTA 1087
DB 1198 TGAGTATATTCATCTGAGAGAAAACCTTACAAATGTGAAAAATGTGAAAAAGCCTTTAA 1257
QY 1088 TAAACA---TCAATCTTTAAGACATCATATATATCTGTTCAATGAAAAAATTAACGTG 1144
DB 1258 TCAATTTCTCACACCTTACTACATTAAGATTAATTCATCTGAGAGAAAACCTTACAAATG 1317
QY 1145 TAAACATGTAAATAAAGTTTTCACCTGACCTTCAAAATTTAGCAACATTAATTAACA 1204
DB 1318 TAAAGATGTGTAAAGCTTTTAAACCTTTCAACCTTCTTAAACATTAAGATTAATTC 1377
QY 1205 TCATGTGATCTCTGCTTATCATGTGATCATCTGTTGTTTAAAAATTTCCAAAC 1264
DB 1378 TACTGAGAGAAAGCTTACAAATCTAAAGATGTGAAAAAGCTTTTAAACCATCTGAAA 1437
QY 1265 TTGGTCAGATTTACATTTTATTAATAACAATGCAATCCAAACTTAAATGTCTTAATG 1324
DB 1438 ACTTACTGAACATTAAGAAAATTCATCTGAGAGAAAACCTTA---TGAATGTGAAAAATG 1494
QY 1325 TGTAAAGGTGTGTGGGAAAAAGTTATCTTCAATATGTAAATGATGATGA 1381
DB 1495 TGGCAAGCTTTTAAACGATCTTCAATCTTACTAGCATTAAGAAAAGTCAATACAGA 1551

RESULT 6
AAS91317
ID AAS91317 standard; cDNA; 4563 BP.
XX
AC AAS91317;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #27121.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN W0200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
XX
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Dmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX
XX P-PSDB; ABG27130.
XX

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 1, SEQ ID NO 27121; 103bp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probe, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 4563 BP; 1752 A; 814 C; 824 G; 1173 T; 0 U; 0 Other;
Query Match 4.6%; Score 95.2; DB 5; Length 4563;
Best Local Similarity 52.7%; Pred. No. 4.2e-08;
Matches 256; Conservative 0; Mismatches 223; Indels 9; Gaps 2;
QY 726 GAAAGTGACAAACCAATGATATCATTTTAAATCTCTCTTCTATCACTGCC 785
DB 1315 GAATGTGGCAAGACCTTTAAACAGTCTCTACACCTTATCAATTAACAAATTCATACT 1374
QY 786 AAAAGTATTTTGCATATGAAAGGTGTGAAGGCCATATAATGACCATCATTTTA 845
DB 1375 GAGAGGAAAACCTTAAAGTGAAGATGTGGCAAGCTTTTAAACATCTTTCAGCCCTT 1434
QY 846 GAGCAACATTTTAAAGAACCCACAGTAATGATGACCGGTATTAATGTACAGTGGACATTTGT 905
DB 1435 ACTTAACATTAAGATTAATTCATCTGAGAGAAAACCATTAAGATGT-----GAAAGATGT 1488
QY 906 GATTAAGCATTTTTCAGAAAATGACATTTGAAAACACATATGTATACATTTCCGAAAAA 965
DB 1489 GGCAAAAGCTTTTGGCAATCTCTACACCTTACTAGACATTAAGCAATTCATATCTGGAGAG 1548
QY 966 AAACCATTCATTTGTCAGTGTGTGTAAAGGGGTAAATCTGACAAACCTTGAAGA 1025
DB 1549 AAACCTTACAAAGTGAAGATGTGGCAAGCTTTTAAACATTTCTGAGACCTTGAAGA 1608
QY 1026 CATGAATCACCCATCAAAAGTCAATTTAAATGTACATTTGAAATTTGTCAAGAGCATTT 1085
DB 1609 CATAGATTAATTCATCTGAGAAAAACCTTACAAATGTGAAGATGTGGAAAAAGCTTTT 1668
QY 1086 TATTAACATCAATCTTTAAG---CATCATATATTTATCTGTTCATGAAAAACATTAACG 1142
DB 1669 AGCCAGTCTCAACCCCTTAAAGAACATCAATATTAATTCATCTGAGAGAAAACCTTACAAA 1728
QY 1143 TGTAAACAATGTAAATTAAGTTTTCACCTGACCTTCAAAATTTAGACAAACATTAATTAATA 1202
DB 1729 TGTGAAGAAATGTGTAAAGCTTTTAAAGTGTCAATCAAAACTTACTGTACATTAAGTAAAT 1788
QY 1203 CATCATGCTG 1212
DB 1789 CATACTGGAG 1798

RESULT 7

AA51569
ID AAL51569 standard; DNA; 2597 BP.
XX
AC AAL51569;
XX
DT 10-APR-2003 (first entry)
XX
DE Human nucleic acid-associated protein coding sequence - SEQ ID NO 52.
XX
KW Human; gene; ds; nucleic acid-associated protein; NAAP; arteriosclerosis;
KW cell proliferative disorder; atherosclerosis; cirrhosis; hepatitis; AIDS;
KW cancer; developmental disorder; renal tubular acidosis; anaemia; asthma;
KW mental retardation; neurological disorder; Alzheimer's disease; epilepsy;
KW Parkinson's disease; autoimmune disorder; inflammatory disorder; allergy;
KW Crohn's disease; transgenic animal; animal model.
XX
OS Homo sapiens.
XX
PN MO2003000864-A2.
XX
PD 03-JAN-2003.
XX
PF 20-JUN-2002; 2002WO-US021179.
XX
PR 22-JUN-2001; 2001US-0300518P.
XX
PR 29-JUN-2001; 2001US-0301787P.
XX
PR 29-JUN-2001; 2001US-0301792P.
XX
PR 29-JUN-2001; 2001US-0301892P.
XX
PR 06-JUL-2001; 2001US-0301893P.
XX
PR 06-JUL-2001; 2001US-0303405P.
XX
PR 15-MAR-2002; 2002US-0364438P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Gandhi AR, Swarnakar A, Hafalia ALA, Warren BA, Emerling BM,
PI Arvizu CS, Ison CH, Honchell CD, Lee EA, Yue H, Forsythe JD;
PI Ramkumar J, Griffin JA, Yang J, Sanjwalala MM, Baughn MR;
PI Borowsky ML, Yao WG, Walla NK, Bandman O, Lal PG, Becha SD, Lee SY;
PI Richardson TW, Elliott VS, Luo W, Tang YT, Zebardjadian Y, Lu Y;
XX
XX WPI; 2003-201420/19.
XX
XX P-PSDB; AAO16419.
XX
DR New nucleic acid-associated proteins and polynucleotides, useful for
XX
XX PT diagnosing, treating or preventing cell proliferative (e.g. cancer),
XX
XX PT neurological (e.g. epilepsy or Parkinson's disease), or autoimmune
XX
XX PT disorders (e.g. AIDS).
XX
XX PS Claim 12; Page 292-293; 312pp; English.
XX
XX CC The invention comprises the amino acid and coding sequences of human
XX
XX CC nucleic acid-associated proteins (NAAP). The DNA and protein sequences of
XX
XX CC the invention are useful for diagnosing, treating or preventing disorders
XX
XX CC associated with aberrant expression of NAAP, such as: cell proliferative
XX
XX CC disorders (e.g. arteriosclerosis, atherosclerosis, cirrhosis, hepatitis
XX
XX CC or cancer); developmental disorders (e.g. renal tubular acidosis, anaemia
XX
XX CC or mental retardation); neurological disorders (e.g. Alzheimer's disease,
XX
XX CC Parkinson's disease or epilepsy); and autoimmune/inflammatory disorders
XX
XX CC (e.g. AIDS, allergies, asthma or Crohn's disease). The DNA sequences of
XX
XX CC the invention are useful for creating transgenic animals to model human
XX
XX CC disease. The present DNA sequence encodes a human nucleic acid-associated
XX
XX CC protein of the invention
XX
SQ Sequence 2597 BP; 947 A; 504 C; 505 G; 641 T; 0 U; 0 Other;
Query March 4.6%; Score 94.6; DB 7; Length 2597;
Best Local Similarity 55.1%; Pred. No. 5.1e-06;
Matches 231; Conservative 0; Mismatches 179; Indels 9; Gaps 2;
QY 797 TTGCACATATGAAAGGTGTGATTAAGCCATATATCAACCATTTATAGACAACTTT 856
DB 1070 TTATTAATGTGAAGATGTGGCAAAAGCCTTTAATGTGTCTCCCAACCTTACTCAACTAA 1129

QY 857 AAGAACCAACAGATATGATGACCCGTATTAATGTACAGTGAACGATTTGTATTAAGCATT 916
DB 1130 GAGAAATTCATCTCGAGAGAAACCTTACAAATGT-----GAAGAGTGTGGCAAAAGCTTT 1183
QY 917 TTTCAAGAAATCATATTTGGAAAACATATATGTATCATATTCGAAAACCAATTCGA 976
DB 1184 TTAACGTCTCTCACTTACTTACATCAATTAAGAAATTTCTACTGGAAAACCATTAACAA 1243
QY 977 TTGTTCAAGTGTGTGTAAGAGGGTTAATTTCTGACAAACCTTGAAGAAACATGAAATCAC 1036
DB 1244 ATGTGAAGATGTGGCAAAAGCCTTTAACAATCTCTCAACCTCAACCAACATTAAGAAAT 1303
QY 1037 CCATTCAAAGTCAATTAATGTATCATTTGAAAATTTGTCAAGAACATTTTATTA--ACA 1093
DB 1304 TCAATCCGAGAGAAACCTTACAAATGTATGAAGATGTGGCAAAAGCCTTTAACCAATTC 1363
QY 1094 TCAATCTTTAAGACATCATATATTAATCTGTCATGAAAAACATTAACGTTAAACAATG 1153
DB 1364 ACACTTCTACATCAATTAAGATTAATTCATCTGAGAGAAACCTTACAAATGTAAAGAAATG 1423
QY 1154 TTAATTAAGTTTCTACTCGACCTTCAAAATTTAGCAACATTAATTAACATCATGTTG 1212
DB 1424 TGGCAAAAGCTTTAAGCGGTCTCAAACTTACTGAACTAGATTAATTCATATCTGGAG 1482
RESULT 8
AAS64586
ID AAS64586 standard; cDNA; 3639 BP.
XX
XX AAS64586;
XX
XX DT 13-FEB-2002 (first entry)
XX
XX DE DNA encoding novel human diagnostic protein #390.
XX
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX
XX KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX OS Homo sapiens.
XX
XX FN WO200175067-A2.
XX
XX PD 11-OCT-2001.
XX
XX PP 30-MAR-2001; 2001WO-US008631.
XX
XX PR 31-MAR-2000; 2000US-00540217.
XX
XX PR 23-AUG-2000; 2000US-00649167.
XX
XX PA (HYSE-) HYSEQ INC.
XX
XX PI Dymnac RT, Liu C, Tang YT;
XX
XX DR WPI; 2001-639362/73.
XX
XX DR P-PSDB; ABG00399.
XX
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX
XX PT responsible for genetic disorders or other traits and to assess
XX
XX PT biodiversity.
XX
XX PS Claim 1; SEQ ID NO 390; 103pp; English.
XX
XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX
XX CC sequences. (I) is useful as hybridisation probes, polymerase chain
XX
XX CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX
XX CC and in recombinant production of (II). The polynucleotides are also used
XX
XX CC in diagnostics as expressed sequence tags for identifying expressed
XX
XX CC genes. (I) is useful in gene therapy techniques to restore normal
XX
XX CC activity of (II) or to treat disease states involving (II). (II) is
XX
XX CC useful for generating antibodies against it, detecting or quantitating a
XX
XX CC polypeptide in tissue, as molecular weight markers and as a food

supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostic, forensic, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAs64197-AAs94564 represent novel human diagnostic coding sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 3639 BP; 1403 A; 708 C; 622 G; 906 T; 0 U; 0 Other;

Query Match 4.5%; Score 92.4; DB 5; Length 3639;

Best Local Similarity 54.1%; Pred. No. 1.3e-07;

Matches 236; Conservative 0; Mismatches 191; Indels 9; Gaps 2;

```
QY 788 AAGTATATTTGACATATGAAAGGCTGATTAAGCTTAATCCGACATCTATTATTA 847
D 738 AAGAGAAACCCCTCAATGGAAGATGCGAAAGCTTTAGCCATTCTCAACCTTGC 797
QY 848 GCACATTTAAGAACCCCAATATGATGACCGCTTAATGATGACAGATTTGA 907
D 798 TAAACATTAAGAAATTCATCTCGAGAAACCCCTCAATGT-----GAAGAAATGTG 851
QY 908 TAAAGCATTTTTCAGAAATACATTTGGAACACATTTGATCTACATTCGAGAAAAA 967
D 852 CAAGCTTTTAGCCGTTCTTCAACCTGCTTAAACATTAAGAAATTCATCTCGAGAA 911
QY 968 ACCATTCCATTGTCAGTGTGTGTAAGGGTAAATTCGACACACCTGAAAAAGA 1027
D 912 ACCCTTCAAAATGTAAGAAATGCGAAAGCTTTTACCAATCTCAACCTTGTCTATCA 971
QY 1028 TGAATACCCCATACAAAGCATTTTAATGTATGTAATGTAATGTAAGAAACATTTTA 1087
D 972 TAGATTAATCTATCTGAGAGAAACCCCTCAATGTAAGAAATGTAAGAAACCTTTAA 1031
QY 1088 TAAACATCAATC---TTTAAACATCATATATATATCTGTCATGAAAAAACATTAACGTG 1144
D 1032 GCGACTCTCAACCTTACTTAACATTAATATATATGCTGAGAGAAACCTTCAAAATG 1091
QY 1145 TAAACATGTAATTAAGTTTTCATCTGACCTTCAAAATTAAGCAACATTAATTAAGA 1204
D 1092 TGAAGATGCGAAAGCTTTTATGATCTTCAATCTTATATACATTAAGTTTATCA 1151
QY 1205 TCATGGTGAATCTCT 1220
D 1152 TACTGAGAGAAACCT 1167
```

RESULT 9

ADB63624

ID ADB63624 standard; cDNA; 2110 BP.

XX ADB63624;

XX 04-DEC-2003 (first entry)

XX Human cDNA encoding clone THWMU20071120.

XX Human; sex: gene; pharmaceutical; diagnostic; gene therapy;

XX tissue regeneration; cell regeneration; membrane protein;

XX signal transduction-related protein; transcription-related protein;

XX osteoporosis; neurological disease; cancer; tumour.

XX Homo sapiens.

XX Key

XX Location/Qualifiers

XX CDS

XX /tag= a

XX /product= "Clone THWMU20071120 protein"

FT

XX EPI308459-A2.
XX 07-MAY-2003.
XX 28-MAR-2002; 2002EP-00007401.
XX 05-NOV-2001; 2001JP-00379298.
XX 25-JAN-2002; 2002US-00350978.
XX (HELI-) HELIX RES INST.
XX (REAS-) RES ASSOC BIOTECHNOLOGY.
XX Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,
XX Yamamoto Y, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I,
XX Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masubo Y;
XX WPI; 2003-450961/43.
XX P-PDB; ADB65594.

New polynucleotides and polypeptides, useful for developing a diagnostic marker or medicines for regulation of their expression and activity, or as targets of gene therapy.

Claim 1; Page; 222pp; English.

The invention discloses a polynucleotide comprising a sequence selected from 1970 fully defined nucleotide sequences which encode novel polypeptides. Also claimed is a polypeptide encoded by the polynucleotide or its partial peptide, an antibody binding to the polypeptide or peptide of the polynucleotide, immunologically assaying the polypeptide or peptide of the polynucleotide by contacting the polypeptide or peptide with the antibody of the encoded protein, and observing the binding between the two, a transformant carrying the polynucleotide in an expressible manner and an antisense polynucleotide. The oligonucleotide is useful as a primer for synthesizing the polynucleotide, or as a probe for detecting the polynucleotide. The polynucleotides and encoded proteins are useful as pharmaceutical agents and many disease-related genes may be included in them, for developing a diagnostic marker or medicines for regulation of their expression and activity, or as targets of gene therapy. The genes are involved in tissue and/or cell regeneration. Membrane proteins, signal transduction-related genes, transcription-related proteins, disease-related proteins and genes encoding them can be used as indicators for diseases (e.g. osteoporosis, neurological diseases, cancer, tumours). The cDNA may be used to regulate the activity or expression of the encoded protein to treat diseases. The sequence presented is a cDNA of the invention. Note: Some of the sequence data for this patent is not represented in the printed specification, but is based on sequence information supplied by the European Patent Office.

Sequence 2110 BP; 784 A; 394 C; 386 G; 546 T; 0 U; 0 Other;

Query Match 4.5%; Score 91.8; DB 9; Length 2110;

Best Local Similarity 51.7%; Pred. No. 1.6e-07;

Matches 314; Conservative 0; Mismatches 277; Indels 16; Gaps 4;

```
QY 780 CGTCCCAAAAAGTATATTTGACATATGAAAGGCTGATTAAGCTTAATCCGACATCA 839
D 1253 CATACGAGAGAAACCCCTCAATGTAAGAAATGCGAAAGCTTTTAAACACTCTCA 1312
QY 840 TTATTAAGCAACATTTTAAGAACCCACAGTAATGATGACCGTAAATGACAGTGCAC 899
D 1313 GCCCTTAATCAATTAAGAAATTCACATCTGAGAGAAACCCCTACAAATGT-----GAA 1366
QY 900 GATTGTATTAAGCATTTTTCAGAAATTCACATTTGGAACACATATTTGATCAATCTCC 959
D 1367 GAATGTGCAAAAGCTTTTAAACGATCTCTCAAAATTTACTGAACATTAAGAACTTCAT 1426
QY 960 GAAAAAAACCATTCATTTGTCAGTGTGTGTAAGGGGTAAATTTCTGACAAACATTTG 1019
D 1427 GAAAGAAACCCCTCAATGTAAGAAATGCGAAAGCTTTTATCAATCTCAAAACCTT 1486
QY 1020 AAAAGCATGAATACCCCATCAAAAGTCAATTTTAATCTAATTTGAAAAATGTCAAGAA 1079
```

the invention relates to a novel human macroprotein-45.76, the encoding polynucleotide, an antagonist and a method of recombinant production. The protein of the invention may be useful for treating dementia, arthralgia, asthma and diabetes. The current sequence is that of the human

PF 28-MAR-2002; 2002EP-00007401.
XX

PR 05-NOV-2001; 2001JP-00379298.
 PR 25-JAN-2002; 2002US-00350978.
 XX
 XX
 PA (HELI-) HELIX RES INST.
 PA (REAS-) RES ASSOC BIOTECHNOLOGY.
 PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I,
 PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y;
 XX
 XX MPI: 2003-450961/43.
 DR P-PSDB; ADB64438.
 XX
 PT New polynucleotides and polypeptides, useful for developing a diagnostic
 PT marker or medicines for regulation of their expression and activity, or
 PT as targets of gene therapy.
 PS
 PS Claim 1; Page; 222pp; English.
 CC The invention discloses a polynucleotide comprising a sequence selected
 CC from 1970 fully defined nucleotide sequences which encode novel
 CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
 CC or its partial peptide, an antibody binding to the polypeptide or peptide
 CC of the polynucleotide, immunologically assaying the polypeptide or
 CC peptide of the polynucleotide by contacting the polypeptide or peptide
 CC with the antibody of the encoded protein, and observing the binding
 CC between the two, a transformant carrying the polynucleotide in an
 CC expressible manner and an antisense polynucleotide. The oligonucleotide
 CC is useful as a primer for synthesizing the polynucleotide, or as a probe
 CC for detecting the polynucleotide. The polynucleotides and encoded
 CC proteins are useful as pharmaceutical agents and many disease-related
 CC genes may be included in them, for developing a diagnostic marker or
 CC medicines for regulation of their expression and activity, or as targets
 CC of gene therapy. The genes are involved in tissue and/or cell
 CC regeneration. Membrane proteins, signal transduction-related proteins,
 CC transcription-related proteins, disease-related proteins and genes
 CC encoding them can be used as indicators for diseases (e.g. osteoporosis,
 CC neurological diseases, cancer, tumours. The cDNA may be used to regulate
 CC the activity or expression of the encoded protein to treat diseases. The
 CC sequence presented is a cDNA of the invention. Note: Some of the sequence
 CC data for this patent is not represented in the printed specification, but
 CC is based on sequence information supplied by the European Patent Office.
 CC
 XX
 SQ Sequence 3078 BP; 1159 A; 516 C; 538 G; 865 T; 0 U; 0 Other;
 Query Match 4.5%; Score 91.8; DB 9; Length 3078;
 Best Local Similarity 52.1%; Pred. No. 1.7e-07;
 Matches 257; Conservative 0; Mismatches 227; Indels 9; Gaps 2;
 QY 730 GTGACGAAACCAATCGATATCATCTTAATATCTTCTTCATCAGTCCCAAA 789
 DB 650 GTGGCAATCATTTTGATGCTTTACACATTAACCTACACATTAAGAAATTCATCTAG 709
 QY 790 AGTATATTGGACATATGAAGGGTGTGATTAAGCCCTAATATGACCATCATTTATTAAGC 849
 DB 710 AGATTCCTTCAATGTGAAGATGTGTAAAGCTTAATGCTCTCAACCCCTTACTA 769
 QY 850 AACATTTAAGAACCCAGATGATGATGACCGCTAATATGACAGTGGAGCATGTGATA 909
 DB 770 AACATTAAGATTAATTCATCTGAGAAAAACCTCAATGT-----GAAGAATGTGCA 823
 QY 910 AAGCATTTTTCAGAAATTCATTTTGAACACATATTTGTATCATTTCCGAAAAAAAC 969
 DB 824 AAGCTTTTAACCGGTCTCTCAATCTTACTTAACATTAATTAATCTATCTGAGAGAAAC 883
 QY 970 CATTCATTTGTCAGTGTGTGTAAGGGGTTAATTCGACCAACACTTGAAGAAACATG 1029
 DB 884 CCTCAATGTGAAGAAATGTGGCAAGCTTTAAACGGTCTCAACCTTACTTAACATA 943
 QY 1030 AATATCCCATCAAAAGTCATTTAATGTACATTTGAAATATGTCAGAAAGACATTTTATA 1089
 DB 944 AAAAATTCATATGAGAAAGAAACCCCTCAATATGTGAAGATGTGGCAAGGCTTTAACC 1003

QY 1090 AACATCAATCTTT---AAGACATCATATATATCTGTTTCANGAAAAAACATTAAGCTGTA 1146
 DB 1004 AGTTCCTCAATTTCTATATTAAGCAATTAAGAAATTCATATGGAATTAACCTCAAAATGTG 1063
 QY 1147 AACAAATGTAATAAAGTTTCACTGACCTTCAAAATGACCAACATTAATTAATAACATC 1206
 DB 1064 AAGAATGTGGCAAGAGCTTTAGAGTATCTCAATTTCTTAATAAACAATTAATATCCATA 1123
 QY 1207 ATGGTGAGTCTCC 1219
 DB 1124 CTGGGAAAAACC 1136
 RESULT 12
 AAD55863
 ID AAD55863 standard; cDNA; 2026 BP.
 XX
 AC AAD55863;
 XX
 DT 07-AUG-2003 (first entry)
 XX
 DE Human nucleic acid associated protein (NAAP)-34 cDNA.
 XX
 XX Human; nucleic acid associated protein; NAAP; stroke; AIDS; neurologic;
 KW cancer; atherosclerosis; neurological; epilepsy; Huntington's disease;
 KW developmental disorder; antiinflammatory; neuroprotective; thymomatous;
 KW Cushing's syndrome; infection; gene therapy; cyostatic; anticonvulsant;
 KW cerebroprotective; hypothyroidism; cell proliferative disorder; allergy;
 KW gene; sr.
 XX
 XX Homo sapiens.
 OS
 FH Key Location/Qualifiers
 FT CDS 126..1973
 FT /tag= a
 FT /product= "Human NAAP protein"
 XX
 PN MO200300648-A2.
 XX
 PD 23-JAN-2003.
 XX
 PF 10-JUL-2002; 2002WC-US021971.
 XX
 PR 12-JUL-2001; 2001US-0305089P.
 PR 12-JUL-2001; 2001US-0305104P.
 PR 13-JUL-2001; 2001US-0305325P.
 PR 13-JUL-2001; 2001US-0305390P.
 PR 19-JUL-2001; 2001US-0306960P.
 PR 20-JUL-2001; 2001US-0306949P.
 PR 27-JUL-2001; 2001US-0308170P.
 XX
 PA (INCYTE-) INCYTE GENOMICS INC.
 XX
 PI Swarnakar A, Richardson TW, Warren BA, Griffin JA, Tang YT;
 PI Yue H, Baughn MK, Emerling BM, Lal PG, Lu DM, Forsythe ID;
 PI Rankumar J, Li JX, Becha SD, Duggan BM, Sanjanwala MM, Lee EA,
 PI Burford N, Elliott VS, Ison CH, Ding L, Borowsky ML, Yao MG;
 PI Barroso I, Tran B, Walla NK, Hafalia AJA, Nguyen DB, Lu Y;
 PI Arvizu CS;
 XX
 XX MPI: 2003-221732/21.
 DR P-PSDB; AAE37047.
 DR
 XX
 XX New human nucleic acid associated proteins (NAAP), useful for diagnosing,
 PT treating and preventing diseases or conditions associated with the
 PT aberrant NAAP expression e.g. cancer, AIDS, atherosclerosis, epilepsy, or
 PT infections.
 PS
 PS Claim 5; Page 259-260; 260pp; English.
 XX
 XX The invention relates to human nucleic acid associated proteins (NAAP)
 CC and their corresponding nucleic acid sequences. The invention is useful
 CC in diagnosing, treating and preventing diseases or conditions associated

CC with the decreased expression or overexpression of NAAp, such as cell
 CC proliferative (e.g. cancer, atherosclerosis), neurological (e.g.
 CC epilepsy, Huntington's disease, stroke), immune/inflammatory (e.g. AIDS,
 CC allergies) and developmental (e.g. Hypothyroidism, Cushing's syndrome)
 CC disorders, or infections. It is also useful in assessing the effects of
 CC exogenous compounds on the expression of nucleic acid and amino acid
 CC sequences of NAAp. The NAAp or its fragments are useful in screening
 CC compounds for which acts as their agonist or antagonist. The microarray
 CC is useful in monitoring or measuring protein-protein interactions, drug-
 CC target interactions, and gene expression profiles. NAAp DNA is used in
 CC gene therapy. The present sequence is human NAAp cDNA
 XX
 SQ Sequence 2026 BP; 699 A; 422 C; 396 G; 509 T; 0 U; 0 Other;

Query Match 4.4%; Score 90.4; DB 7; Length 2026;
 Best Local Similarity 51.8%; Pred. No. 2.9e-07;

Matches 258; Conservative 0; Mismatches 231; Indels 9; Gaps 2;

QY 726 GAAAGTGAAGAAACCAATGCATATCATCTTTAAATCTCTTCTTCTTCAATCACTGCC 785
 DB 1002 GAATGTGGCAAGCCCTTCAATGCTTCAACCTTCAACATCAAGTAAATTCATCT 1061
 QY 786 AAAAGTATATTTGCAATATGAAGGTGTGAATGAAGCCATATGACATCATATTA 845
 DB 1062 GGAGAGAAAGCCCTTCAATGTGAAGATGTGTAAGCTTTAAACCACTTCAAGCCTT 1121
 QY 846 GAGCAACATTTAAGAACCCACAGTAATGATCAAGGTATAATGTACAGTGAAGATGT 905
 DB 1122 ACTACACATMAAGTTCATTCATGTATGAAGAAAACCTCAAAATGT-----GAAAGATGT 1175
 QY 906 GATAAAGCATTTTTCAGAAATATCATCTTTGAAACATATGTATCATCTCCGAAAA 965
 DB 1176 GACAAAGCTTTTAAACGATTCCTATCTTAAACATMAAGTATTCATTCCTGGAGAG 1235
 QY 966 AAACATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 1025
 DB 1236 AAATCTCAAAATGTGAACATGTGGCAAGGCTTTAACTGTCTTCAACCTTCAAAA 1295
 QY 1026 CATGAATATCCCATTAACAAAGTCATTTAAATGTACATTTGAAATTTGCAAGACATTT 1085
 DB 1296 CATGAAGAAATTCATCACTGAGAGAAACCTTCAAAATGTGAAGATGTGGCAAGCCTTT 1355
 QY 1086 TATAAATCAT---CAATCTTTAAGACATCATATATATATCTGTCATGAAAAAATTAACG 1142
 DB 1356 AATGTCTTCAACCTTCACTACATTAAGATTCATCACTGAGAGAAACCTTCAAAA 1415
 QY 1143 TGTAAACATGTATTAAGTTTCACTGACCTTCAAAATTAAGCACAACATTAATTAAA 1202
 DB 1416 TGTGAAGATGTGGCAAGCCTTTAAACCACTCTCAAAATTAATCTATTCATTAAGTAAAT 1475
 QY 1203 CATCATGTGGATCTCCT 1220
 DB 1476 CATACTGGAGAAACCT 1493

RESULT 13

ADA53124
 ID ADA53124 standard; cDNA; 2114 BP.

AC ADA53124;

DT 20-NOV-2003 (first entry)

DE Human coding sequence, SEQ ID 692.

XX Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;

KM Gene Therapy; human; secretory protein; membrane proteins; cancer;

KM inflammatory disease; osteoporosis; neurological disease; gene; ss.

OS Homo sapiens.

XX EPI293569-A2.

FN

XX

PD 19-MAR-2003.
 XX
 PF 21-MAR-2002; 2002EP-0006586.
 XX
 PR 14-SEP-2001; 2001IP-00328381.
 PR 24-JAN-2002; 2002US-0350435P.
 XX
 PA (HELI-) HELIX RES INST.
 PA (REAS-) RES ASSOC BIOTECHNOLOGY.
 XX
 PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I,
 PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y,
 DR P-PDB; ADA54763.
 XX
 DR MPI; 2003-395539/38.

PT New polynucleotides encoding full-length polypeptides, e.g. secretory
 PT and/or membrane proteins, useful for developing medicines for diseases in
 PT which the gene is involved, or as target molecules for gene therapy.

PS Claim 1; SEQ ID NO 692; 205pp; English.

CC The present invention relates to novel human secretory or membrane
 CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
 CC ADA54071). The coding sequences are useful in the gene therapy of
 CC diseases caused by abnormalities of the proteins, e.g. cancer,
 CC inflammatory diseases, osteoporosis or neurological disease.

SQ Sequence 2114 BP; 717 A; 439 C; 434 G; 524 T; 0 U; 0 Other;

Query Match 4.4%; Score 90.4; DB 7; Length 2114;
 Best Local Similarity 51.8%; Pred. No. 2.9e-07;

Matches 258; Conservative 0; Mismatches 231; Indels 9; Gaps 2;

QY 726 GAAAGTGAAGAAACCAATGCATATCATCTTTAAATCTCTTCTTCTTCAATCACTGCC 785
 DB 1024 GAATGTGGCAAGCCCTTCAATGCTTCAACCTTCAACATCAAGTAAATTCATCT 1083
 QY 786 AAAAGTATATTTGCAATATGAAGGTGTGAATGAAGCCATATGACATCATATTA 845
 DB 1084 GAGAGAAAGCCCTTCAAAATGTGAAGATGTGTGAAGCTTTAAACCACTTCAAGCCTT 1143
 QY 846 GAGCAACATTTAAGAACCCACAGTATGATGATGATTAATGTACAGTGGAGATGT 905
 DB 1144 ACTACACATTAAGTTCATTCATGTTAAAGAAAACCTTCAAAATGT-----GAAAGATGT 1197
 QY 906 GATAAAGCATTTTTCAGAAATATCATCTTTGAAACATATGTATCATCTCCGAAAA 965
 DB 1198 GACAAAGCTTTTAAACGATTCCTATCTTAAACATTAAGTATTCATTCCTGGAGAG 1257
 QY 966 AAACATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 1025
 DB 1258 AAATCTCAAAATGTGAACATGTGGCAAGGCTTTAACTGTCTTCAACCTTCAAAA 1317
 QY 1026 CATGAATATCCCATTAACAAAGTCATTTAAATGTACATTTGAAATTTGCAAGACATTT 1085
 DB 1318 CATGAAGAAATTCATTCATGAGAGAAACCTTCAAAATGTGAAGATGTGGCAAGCCTTT 1377
 QY 1086 TATAAATCAT---CAATCTTTAAGACATCATATATATATCTGTCATGAAAAAATTAACG 1142
 DB 1378 AATGTCTTCAACCTTCACTACATTAAGATGATTCATCACTGAGAGAAACCTTCAAAA 1437
 QY 1143 TGTAAACATGTATTAAGTTTCACTGACCTTCAAAATTAAGCACAACATTAATTAAA 1202
 DB 1438 TGTGAAGATGTGGCAAGCCTTTAAACCACTCTCAAAATTAATCTATTCATTAAGTAAAT 1497
 QY 1203 CATCATGTGGATCTCCT 1220
 DB 1498 CATACTGGAGAAACCT 1515

RESULT 14

ADA52931
ID ADA52931 standard; cDNA; 2230 BP.
XX
AC ADA52931;
XX
XX 20-NOV-2003 (first entry)
XX
DE Human coding sequence, SEQ ID 499.
XX
KM Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
XX Gene Therapy; human; secretory protein; membrane proteins; cancer;
XX inflammatory disease; osteoporosis; neurological disease; gene; ss.
XX Homo sapiens.
OS
XX EPI293569-A2.
XX
XX 19-MAR-2003.
XX
XX 21-MAR-2002; 2002EP-0006586.
XX
XX 14-SEP-2001; 2001UP-0032831.
XX
XX 24-JAN-2002; 2002US-0350435P.
XX
XX (HELI-) HELIX RES INST.
XX (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
XX Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
XX Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
XX Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y;
XX WPI; 2003-395539/38.
XX P-PSDB; ADA54570.
XX
XX New polynucleotides encoding full-length polypeptides, e.g. secretory
XX and/or membrane proteins, useful for developing medicines for diseases in
XX PT which the gene is involved, or as target molecules for gene therapy.
XX
XX Claim 1; SEQ ID NO 499; 205pp; English.
XX
XX The present invention relates to novel human secretory or membrane
XX CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
XX CC ADA54071). The coding sequences are useful in the gene therapy of
XX CC diseases caused by abnormalities of the proteins, e.g. cancer,
XX CC inflammatory diseases, osteoporosis or neurological disease.
XX
XX Sequence 2230 BP; 831 A; 434 C; 400 G; 565 T; 0 U; 0 Other;
SQ
Query Match 4.4%; Score 90.2; DB 7; Length 2230;
Best Local Similarity 50.1%; Pred. No. 3.2e-07;
Matches 338; Conservative 0; Mismatches 323; Indels 14; Gaps 4;
QY 710 ACTTATTCATGATGAGAAAGTACGAACCAATGCATATCATCTTAAATCTTCTTC 769
Db 640 ACATTTAGATGATGAAGATGTGCAATACCTTGCATGCTTCAAGCCTAATCAACA 699
QY 770 TTCTTCATCAGTCCCAAAAGTATTTGCAATATGAAGGGTGTGATTAAGCTTAA 823
Db 700 TAAAAAATTCATCTAGAGAGAAATTTCTACAAATGTGAAGAGTGGAAAAACCTTAA 759
QY 830 TCGACCATCATTTATAGAGCAACATTTAAGAACCCAGATGATGACCGCTTAATAG 889
Db 760 CTGGTCCACAAACCTTTCTAAACCTAAGAAATTCATCTGAGAAAAAACCTTCAATG 819
QY 890 TACAGTGACGATTTGTGATTAAGCATTTTTCAGAAAAATCATTTGGAAACACATATGT 949
Db 820 TGAAGTA-----TGTGAAAAGCCTTTCACCAATTCCTCATCTTACTAAACATAAGAT 873
QY 950 ATCACATTCGAAAAAACCATTCCATGTTCTAGTGTGTGTAAGGGGTTAATTCGCG 1009
Db 874 AATTCTGACTGGAAGAAAAACCTTAAATGTGCACTGTGTGCAAAAGCCTTTAAACAGTC 933
QY 1010 ACAACATTTGAAAAGACATGAATATCACCATACAAAGTCATTTAAATGATCATTTGAAA 1069

Db 934 CTCACACTTACTAGACATTAAGATTAATTCATATCTGTAAGAGAAACCTACAAATGTGAACA 993
QY 1070 TTGTCAAGAAAGCATTTTATTAACATCAATCTTTA--AGACATCATATATTCGTGTTCA 1126
Db 994 ATGTGGCAAGGCTTTTAAGCAGTCCCAACCTTCAATCAACATCAAGATTAATTAATCTGG 1053
QY 1127 TGAAAAAACATTAAGGTGTAACATATGATTAATTAAGTTTCACTGACCTTCAAAATTAGC 1186
Db 1054 AAGAGAACCATACAAATGTGAGAAATGTGGCAAAAGCTTTTAACATATCTTAACAACTTAC 1113
QY 1187 ACAACATTAATTAACAAATCATATGATG- GATCTCGCTTATCAATGTGATTCATCTGGTT 1245
Db 1114 TGAACATTAAGAAATTTACATAGAGAGAAACCTTCAAAATGTGAAGATGTGGCAAAAGC 1173
QY 1246 GTTTTAAAAATTTCCAAACTTGTCAGATTTCAATTTCAATTAACAAAGCATGCAATCCAA 1305
Db 1174 CTTTAAACGATTTTCAACCTTATTTACATTAAGATTAATTCATAGCAGAGAGAAACCC-- 1231
QY 1306 AACTTAATATGCTCAATATGTGTAAGGTTGTTGGAAAAAAGGTTTATCTTCACATA 1365
Db 1232 --CACAAATGTGAAGAAATGTGGCAGAGCTTTTAAACCAATGCCGAAGCTCACTGAACATA 1289
QY 1366 TGTTAAGTCATGATG 1380
Db 1290 AGTTAATTCATCTG 1304
RESULT 15
ADC58104
ID ADC58104 standard; cDNA; 1952 BP.
XX
XX ADC58104;
AC
XX 18-DEC-2003 (first entry)
DT
XX Zinc finger protein 85-54.12 cDNA SEQ ID NO:1.
XX
XX ss; gene; zinc finger protein; cancer; haemopathy; HIV; immune disease.
XX
XX Unidentified.
OS
XX
XX Key Location/Qualifiers
FH 173.1651
FT CDS /*tag= a
FT /product= "zinc finger protein 85-54.12"
XX
XX CN1382706-A.
PN
XX
XX 04-DEC-2002.
PD
XX 26-APR-2001; 2001CN-00112739.
PF
XX 26-APR-2001; 2001CN-00112739.
PR
XX
XX (BIOW-) BIOWINDOM GENE DEV INC SHANGHAI.
PA
XX
XX Mao Y, Xie Y;
PI
XX
XX WPI; 2003-269470/27.
DR P-PSDB; ADC58105.
XX
XX Polypeptide-zinc finger protein 85-54.12 and polynucleotide for coding
PT it.
XX
XX Claim 6; SEQ ID NO 1; 33pp; Chinese.
XX
XX The invention relates to a novel zinc finger protein 85-54.12. The
XX CC polynucleotide encoding the zinc finger protein, the application of the
XX CC protein in treating diseases such as cancer, haemopathy, HIV infection
XX CC and immune disease, the antagonist of the polypeptide and its medical
XX CC action, and the application of the polynucleotide are also disclosed. The
XX CC present sequence encodes the zinc finger protein of the invention.

XX Sequence 1952 BP; 760 A; 386 C; 321 G; 485 T; 0 U; 0 Other;

Query Match 4.4%; Score 90; DB 9; Length 1952;
Best Local Similarity 52.8%; Pred. No. 3,4e-07;
Matches 267; Conservative 0; Mismatches 230; Indels 9; Gaps 3;

```

QY 710 ACTTTATTCATGAGGAAAGTGCAGAAACCAATGCATATCATCTTTAATATCTCTTC 769
Db 469 ACCTTTAAATATTAAGAAATTTGGCAATCATTTTGCATATTTTCAAACTTAACCTCAACA 528
QY 770 TTCTTCATCAGTCCCAAAAAGTATTTGACATATGAAGGGGTGAATTAAGCTATTA 829
Db 529 TAAATAATTTGTACTAGAGTAAATTTCTACAAATGTGAAGACTGTGAAAGAGCTTTTAA 588
QY 830 TCGACCATGATTAATTAGAGCAACATTTAAGAACCAAGTAATATCGAACCATAAATG 889
Db 589 TGGATCCTCAATCTTACTTAACATTAAGAAATTCATATTGGAGAGAAATCGTACATATG 648
QY 890 TACAGTGAAGATTGTGATTAAGCATTTTCAGAAATCACATTGSAACACATATTTGT 949
Db 649 T-----GAGAAATGTGGCAAGCCTGTAAACAGTTCACAAACTTACTACATTAAGAT 702
QY 950 ATCACATTCGAAAAAACAATTCATGTTCAGTGTGGTAAAGGGTTAATTCG 1009
Db 703 AATTATTACTAGAGCAAACTCTACAAAGTGAAGAAATGAGCAAGCCTTAAACCTGTC 762
QY 1010 ACACACCTTGAAGAAGCATGAATCAACCATACAAAGTCATTAAATGTACATTGAAAA 1069
Db 763 GTACACATTCACACACATACATATATCATCTGAGAGAAATCCCTACAAACGTGAAGA 822
QY 1070 TTGTCAAGAGCATTTTAT-AAAACATCATCTTTAAGACATCATATATATCTGTTCATG 1128
Db 823 ATGTACAAAGGCTTTTAAACGCTCTCAACCTTACTACACATTAAGATATATTCATACAG 882
QY 1129 AAAAAACATTAAAG--TGTAACAATGTAAATTAAGTTTCACTCGACCTCAAAATTAAGC 1186
Db 883 AGAGAACTCAATGAATTAAGAAATGTGGCAAGCTTTCAACGATCCCAACACCTTAC 942
QY 1187 ACAACATTAATTAATAACATCATGTG 1212
Db 943 CAGACATTAAGATTAATTCATCTGAG 968

```

Search completed: July 28, 2004, 15:04:04
Job time : 828 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 28, 2004, 12:51:57 ; Search time 8116 Seconds

(without alignments)
11001.312 Million cell updates/sec

Title: US-09-831-804-1

Perfect score: 2060

Sequence: 1 cttattaggaagatgctc.....gacatactactcttcatcg 2060

Scoring table:

IDENTITY_NDC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rtd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|-------------|
| 1 | 2060 | 100.0 | 2060 | 6 | BD274349 |
| 2 | 1239 | 60.1 | 1239 | 6 | BD274350 |
| 3 | 1235.8 | 60.0 | 1239 | 6 | AX489193 |
| 4 | 217.4 | 10.6 | 1560 | 8 | YSCNOFAT |
| 5 | 217.4 | 10.6 | 1739 | 8 | YSCFPIIA |
| 6 | 217.4 | 10.6 | 37497 | 8 | YSCP9677 |
| 7 | 199.2 | 9.7 | 2133 | 8 | YSCRP026A |
| 8 | 193 | 9.4 | 797 | 11 | CNSO61NG |
| 9 | 101 | 4.9 | 169063 | 9 | AC138126 |
| 10 | 101 | 4.9 | 177299 | 9 | AC073544 |
| 11 | 99.4 | 4.8 | 2132 | 6 | AX714042 |
| 12 | 99.4 | 4.8 | 2132 | 6 | AX056088 |
| 13 | 99.4 | 4.8 | 138627 | 9 | AC099500 |
| 14 | 99.4 | 4.8 | 169500 | 2 | AC138469 |
| 15 | 98.2 | 4.8 | 2237 | 9 | AK122869 |
| 16 | 97.8 | 4.7 | 1203 | 9 | D70831 |
| 17 | 97.8 | 4.7 | 235332 | 9 | AC008739 |
| 18 | 97.2 | 4.7 | 115995 | 9 | AC011494 |
| 19 | 96.6 | 4.7 | 2320 | 6 | AR270491 |
| 20 | 96.6 | 4.7 | 2320 | 9 | HS035376 |
| 21 | 96.6 | 4.7 | 68304 | 10 | BX000432 |
| 22 | 95.2 | 4.6 | 1389 | 9 | M27879 |
| 23 | 95 | 4.6 | 981 | 9 | HS42F9 |
| 24 | 94.6 | 4.6 | 186233 | 9 | AC092329 |
| 25 | 94.6 | 4.6 | 189317 | 2 | AC024483 |
| 26 | 94.2 | 4.6 | 292390 | 2 | AC105677 |
| 27 | 93.6 | 4.5 | 2873 | 9 | HDMHPLK |
| 28 | 93.6 | 4.5 | 2873 | 11 | G28705 |
| 29 | 93.6 | 4.5 | 2873 | 11 | HMSMS1269 |
| 30 | 93.6 | 4.5 | 155439 | 9 | AC008626 |
| 31 | 93.6 | 4.5 | 191082 | 2 | AC141066 |
| 32 | 93.6 | 4.5 | 199585 | 10 | AC124426 |
| 33 | 93.6 | 4.5 | 203396 | 9 | AC073210 |
| 34 | 92.4 | 4.5 | 41153 | 9 | AC016628 |
| 35 | 92.4 | 4.5 | 158430 | 2 | AC012431 |
| 36 | 92 | 4.5 | 156835 | 9 | AC010620 |
| 37 | 91.8 | 4.5 | 1870 | 9 | AY044432 |
| 38 | 91.8 | 4.5 | 2110 | 6 | AX748253 |
| 39 | 91.8 | 4.5 | 2110 | 6 | AX748253 |
| 40 | 91.8 | 4.5 | 2909 | 9 | AK093669 |
| 41 | 91.8 | 4.5 | 2976 | 9 | HSMB808021 |
| 42 | 91.8 | 4.5 | 3078 | 6 | AX747097 |
| 43 | 91.8 | 4.5 | 3078 | 9 | AK091618 |
| 44 | 91.8 | 4.5 | 3123 | 9 | BC040594 |
| 45 | 91.8 | 4.5 | 174994 | 9 | AC022145 |

ALIGNMENTS

RESULT 1
BD274349
LOCUS
DEFINITION
Candida albicans tflr1 gene (Catf1r1) and the coded CATF1R1 protein.
ACCESSION
BD274349
VERSION
BD274349.1 GI:33084117
KEYWORDS
JP 2002531068-A/1.
SOURCE
Candida albicans
ORGANISM
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Mitosporic Saccharomycetales; Candida.
REFERENCE
1 (bases 1 to 2060)
Pallier, F.B., Camier, S. and Senne, A.
TITLE
Candida albicans tflr1 gene (Catf1r1) and the coded CATF1R1

JOURNAL
Patent: JP 2002531068-A 1 24-SEP-2002;
AVENTIS PHARMA SA
Candida albicans
OS JP 2002531068-A/1
PN 24-SEP-2002
PD 09-NOV-1999 JP 2000581204
PR 10-NOV-1998 FR 98/14147
PI FLORENCE BORDON PALLIER, SYLVIE CAMIER, ANDRE SENTENAC PC
C12N15/09, C12N15/09, A61K39/00, A61K45/00, A61P31/10, C07K14/40, PC
C07K16/14,
PC C12N1/19, C12N1/21, C12P21/02, C12Q1/02, G01N33/15, G01N33/50, PC
G01N33/53,
PC G01N33/566, G01N33/569// (C12N15/09, C12R1:725), C12N15/00, C12N15/00,
PC 00,
PC (C12N15/00, C12R1:725)
CC Candida albicans effilagine (CatfIIIA) and the coded CATFIIIA
CC protein
FH Key
FT source
1. .2060
Location/Qualifiers
/organism='Candida albicans',
/mol_type='genomic DNA'
/DB_xref='taxon:5476'

ORIGIN
Query Match 100.0%; Score 2060; DB 6; Length 2060;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2060; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTATTAGAGAGATTGGCTAGAGCCATTGTTATACGGGCTTCGCAAGTGCATTTGTTT 60
DB 1 CTTTATTAGAGAGATTGGCTAGAGCCATTGTTATACGGGCTTCGCAAGTGCATTTGTTT 60
QY 61 TACTAAATATCCCAATCATTTGGGCTTCAGTGAATGGGGGTGTCATCTCTGGTGTAG 120
DB 61 TACTAAATATCCCAATCATTTGGGCTTCAGTGAATGGGGGTGTCATCTCTGGTGTAG 120
QY 121 AAATAGGCGAGGCGCTCCGATCCCAAAAAAGAAAGATCAGATGTCCTGGCTGCAGA 180
DB 121 AAATAGGCGAGGCGCTCCGATCCCAAAAAAGAAAGATCAGATGTCCTGGCTGCAGA 180
QY 181 TTGTAGCCATGCGAAATGCCGAAAAATGAAAAAAGAAAAAAGTCTACTGGGCCACT 240
DB 181 TTGTAGCCATGCGAAATGCCGAAAAATGAAAAAAGAAAAAAGTCTACTGGGCCACT 240
QY 241 ACAAAAGAAAAAGATTGAACATGATCAGATGATGCTGGAGCCCTCATATATTATTA 300
DB 241 ACAAAAGAAAAAGATTGAACATGATCAGATGATGCTGGAGCCCTCATATATTATTA 300
QY 301 TATGTGACGGGCTTTAGAAATTTGTATATTTGTGTCTGACACTGTGTGTTAATCT 360
DB 301 TATGTGACGGGCTTTAGAAATTTGTATATTTGTGTCTGACACTGTGTGTTAATCT 360
QY 361 GGACATCTGTTCCCTTGTGAAGGTCGTCTGTAATGAATTCATGATCAAGATTAAT 420
DB 361 GGACATCTGTTCCCTTGTGAAGGTCGTCTGTAATGAATTCATGATCAAGATTAAT 420
QY 421 GACTTGTCTACCTCATAGATGCGGACTTGATTAATTAATGAGCTTTAATCCTCTGATA 480
DB 421 GACTTGTCTACCTCATAGATGCGGACTTGATTAATTAATGAGCTTTAATCCTCTGATA 480
QY 481 TATGTGAACCACTTGACTTATTTCTTGTGTGGGATTCACCTTGATGATGATGTTAAG 540
DB 481 TATGTGAACCACTTGACTTATTTCTTGTGTGGGATTCACCTTGATGATGATGTTAAG 540
QY 541 CAATGTATTTGTCATCTTTTGTCTTGTGCGCACTTCCTTAAATTCGGAAT 600
DB 541 CAATGTATTTGTCATCTTTTGTCTTGTGCGCACTTCCTTAAATTCGGAAT 600
QY 601 ATTTCATTAATGAGCGAGCGACTTCTCTCTCATAGAAAAAATAAATAAATACTGA 660
DB 601 ATTTCATTAATGAGCGAGCGACTTCTCTCTCATAGAAAAAATAAATAAATACTGA 660

DB 601 ATTTCATTAATGAGCGAGCGACTTCTCTCTCATAGAAAAAATAAATAAATACTGA 660
QY 661 AAAAATTAACACGGGACCTCATCTCTTTTCAAAATCACCTTTATTAATTAATCA 720
DB 661 AAAAATTAACACGGGACCTCATCTCTTTTCAAAATCACCTTTATTAATTAATCA 720
QY 721 TGAATGAAAGTACGCAAAACCAATGATCATCTTTAATATCTTCTCTTCATCAG 780
DB 721 TGAATGAAAGTACGCAAAACCAATGATCATCTTTAATATCTTCTCTTCATCAG 780
QY 781 GTCCCAAAAATATTTTGCACATATGAAAGGTCGTATTAAGCCCTATATGCAACATCA 840
DB 781 GTCCCAAAAATATTTTGCACATATGAAAGGTCGTATTAAGCCCTATATGCAACATCA 840
QY 841 TATTAGAGCAATTTTAAGACCCAGATTAATGATGACCGATTAATGATGAGAGAG 900
DB 841 TATTAGAGCAATTTTAAGACCCAGATTAATGATGACCGATTAATGATGAGAGAG 900
QY 901 ATTGTGATTAAGCAATTTTCAAGAAATCATTGGAACACATATTTGATCAGATTCCG 960
DB 901 ATTGTGATTAAGCAATTTTCAAGAAATCATTGGAACACATATTTGATCAGATTCCG 960
QY 961 AAAAAAACCATTCATTTGTTCACTGTGTGTGAAGGGGTAAATTCGCAACACTTGA 1020
DB 961 AAAAAAACCATTCATTTGTTCACTGTGTGTGAAGGGGTAAATTCGCAACACTTGA 1020
QY 1021 AAAGACATGAATTCACCCATACAAAGTCATTTAAATGATGATGAAATTTGCAAGAG 1080
DB 1021 AAAGACATGAATTCACCCATACAAAGTCATTTAAATGATGATGAAATTTGCAAGAG 1080
QY 1081 CATTATTAATCAATCATCTTTAAGACATCATATATATATCTGTTCAAGAAAAATTA 1140
DB 1081 CATTATTAATCAATCATCTTTAAGACATCATATATATCTGTTCAAGAAAAATTA 1140
QY 1141 CGTGAACATGTAATTAAGATTTCATCTGACCTCAAAATTAAGACATTAATTA 1200
DB 1141 CGTGAACATGTAATTAAGATTTCATCTGACCTCAAAATTAAGACATTAATTA 1200
QY 1201 AACATCATGATGATCCTCGCTTATCAATGATCATCCTGCTGTTTAAAAATTTCC 1260
DB 1201 AACATCATGATGATCCTCGCTTATCAATGATCATCCTGCTGTTTAAAAATTTCC 1260
QY 1261 AAACCTGTCAGTATTAACAATTTCAATTAATAACAATGATCCAAATCTTAATGTCTTA 1320
DB 1261 AAACCTGTCAGTATTAACAATTTCAATTAATAACAATGATCCAAATCTTAATGTCTTA 1320
QY 1321 AATGTGTAAAGTGTGTGGTGGAAAAAGGTTATCTTCAATATGTAATGATGATG 1380
DB 1321 AATGTGTAAAGTGTGTGGTGGAAAAAGGTTATCTTCAATATGTAATGATGATG 1380
QY 1381 ATTCTACATGATCAAAATATGACCTTGATTAATGATGATGATGATGATGATGATG 1440
DB 1381 ATTCTACATGATCAAAATATGACCTTGATTAATGATGATGATGATGATGATGATG 1440
QY 1441 AAAATGAATTAAGTTGAACATTTAATATCTTCAATGATGATGATGATGATGATGAT 1500
DB 1441 AAAATGAATTAAGTTGAACATTTAATATCTTCAATGATGATGATGATGATGATGAT 1500
QY 1501 TAAAGGAAATGAAATTAAGAAACCTTAATGATCAAGATGCAATTAATTAATA 1560
DB 1501 TAAAGGAAATGAAATTAAGAAACCTTAATGATCAAGATGCAATTAATTAATA 1560
QY 1561 ATTGTGATGAATTTGAACACAGAGAAATTAAGTGAAGAAAGATGAAGAGATGAAGAG 1620
DB 1561 ATTGTGATGAATTTGAACACAGAGAAATTAAGTGAAGAAAGATGAAGAGATGAAGAG 1620
QY 1621 ATAGCTATGATGAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATG 1680
DB 1621 ATAGCTATGATGAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATG 1680
QY 1681 TAAATCATTTACTGCTCTTCTTGAAGAGTTCAGAGAGGTTCTAATCTTATTTGAATA 1740
DB 1681 TAAATCATTTACTGCTCTTCTTGAAGAGTTCAGAGAGGTTCTAATCTTATTTGAATA 1740

| | | | |
|----|------|--|------|
| QY | 1741 | GTGGGAAGAAGTCAATGTCTCTTAAGATTAATGTGATAGATGTTTCTAGAGAAATAG | 1800 |
| Db | 1741 | GTGGGAAGAAGTCAATGTCTCTTAAGATTAATGTGATAGATGTTTCTAGAGAAATAG | 1800 |
| QY | 1801 | ATTACGTGACATTGGAATGGCATGATGTAATTTACAAGAATTAAGTCATTTCTTAA | 1860 |
| Db | 1801 | ATTACGTGACATTGGAATGGCATGATGTAATTTACAAGAATTAAGTCATTTCTTAA | 1860 |
| QY | 1861 | ATAGATAGAAAAAGAGAATCCAGAAAGTGAAACAATGGTATAAAAAACCCAGATGG | 1920 |
| Db | 1861 | ATAGATAGAAAAAGAGAATCCAGAAAGTGAAACAATGGTATAAAAAACCCAGATGG | 1920 |
| QY | 1921 | ATTATATGCCAATAGAAACATCAGTGAATTTCTCATATATATACATTTAAAAATTATATA | 1980 |
| Db | 1921 | ATTATATGCCAATAGAAACATCAGTGAATTTCTCATATATATACATTTAAAAATTATATA | 1980 |
| QY | 1981 | CATTTTATATTCCTTTAATTTTAAATTTTGTGGGCTTTTATTTACATTAATTAAGTT | 2040 |
| Db | 1981 | CATTTTATATTCCTTTAATTTTAAATTTTGTGGGCTTTTATTTACATTAATTAAGTT | 2040 |
| QY | 2041 | GACATATTACCTCTTAATG | 2060 |
| Db | 2041 | GACATATTACCTCTTAATG | 2060 |

| RESULT 2 | LOCUS | DEFINITION |
|----------|----------|--|
| BD274350 | BD274350 | 1239 bp DNA linear PAT 17-JUL-2003 |
| | | <i>Candida albicans</i> tfl1 gene (catfl1) and the coded Catfl1A |

| | |
|-----------|--------------------|
| ACCESSION | BD274350 |
| VERSION | BD274350.1 |
| KEYWORDS | JP 2002531068-A/2. |
| SOURCE | Candida albicans |
| ORGANISM | Candida albicans |

REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 1239)
Pallier,F.B., Camter,S. and Sentenac,A.
Candida albicans tllflingene (Catfllia) and the coded CATPflTA
Patent: JP 2002531068-A 2 24-SEP-2002;
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes
Saccharomycetales; mitosporic Saccharomycetales; Candida.

| COMMENT | OS | PN | PD |
|---------|------------------|------------------|-------------|
| | Candida albicans | JP 2002531068-A/ | 24-SEP-2002 |

| Accession | Protein | Gene | Organism | Length (aa) | Weight (kDa) | PI | Ref |
|---------------|---------------------|------------------|---------------|---------------|--------------|-----------|----------------|
| PF000581.204 | | | | | | | 09-NOV-1999 JP |
| FR10-NOV-1998 | | | | | | | FR 98/14147 |
| P1 | FLORENCE BORDON | PALLITER, SYLVIE | CAMIER, ANDRE | SENTENAC | PC | | |
| C12N15/09 | C12N15/09 | A6IK3/00 | A6IK45/00 | A6IP31/10 | C07K14/40 | PC | |
| C07K16/14 | | | | | | | |
| PC | C12N1/19 | C12N1/21 | C12P21/02 | C13Q1/02 | G01N33/15 | G01N33/50 | PC |
| G01N33/53 | | | | | | | |
| PC | G01N33/566 | G01N33/569 | | (C12N15/09 | C12R1:725) | C12N15/00 | C12N15/00 |
| PC | (C12N15/00 | C12R1:725) | | | | | |
| CC | Candida albicans | tfIIIA gene | (CafIIIA) | and the coded | CafIIIA | | |
| CC | protein | | | | | | |
| Key | | | | | | | |
| FM | Location/Qualifiers | | | | | | |
| CDS | (1) . . (1236) | | | | | | |

| FEATURES | Location/Qualifiers |
|----------|---------------------|
| source | 1..1239 |

```

/organism="Candida albicans"
/mol_type="genomic DNA"
/db_xref="taxon:5476"

```

| | | | | |
|-----------------------|-----------------|---------------------|-----------|--------------|
| Query Match | 60.1%; | Score 1239; | DB 6; | Length 1239; |
| Best Local Similarity | 100.0%; | Pred. No. 2.6e-187; | | |
| Matches 1239; | Conservative 0; | Mismatches 0; | Indels 0; | Gaps 0; |

QY 720 ATGAGTGAAGCTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTGATCA 779
|||||

| | | | |
|----|------|---|------|
| Db | 1 | ATGAGTGAAGACGAGAAAACCAATCGATATCATCTTTAATATCTTCTCTCTTCATCA | 60 |
| Qy | 780 | CGTCCCAAAAAGTATATTTTGGACATATGAAAGGTTGATTAAGCCCTAATAGCACATCA | 839 |
| Db | 61 | CGTCCCAAAAAGTATATTTTGGACATATGAAAGGTTGATTAAGCCCTAATATAGCACATCA | 120 |
| Qy | 840 | TTATTTAGGCAACATTTTAAGAACCCACAGTAATGATCGACGGTAAATGTACAGTGCAC | 899 |
| Db | 121 | TTATTTAGGCAACATTTTAAGAACCCACAGTAATGATCGACGGTAAATGTACAGTGCAC | 180 |
| Qy | 900 | GATTGTGATPAAAGCATTTTTCAGAAAATCACATTTGGAAACACATTTGTATCATATTC | 959 |
| Db | 181 | GATTGTGATPAAAGCATTTTTCAGAAAATCACATTTGGAAACACATTTGTATCATATTC | 240 |
| Qy | 960 | GAATAAAAAACCATTCATTTGTTCAAGTGTGGTAAAGGGGTAAATTCCTGCACACCTTG | 101 |
| Db | 241 | GAATAAAAAACCATTCATTTGTTCAAGTGTGGTAAAGGGGTAAATTCCTGCACACCTTG | 300 |
| Qy | 1020 | AAAAACATGAAATACCCCATACAAAGTCATTTAAATGTACATTTGAAATTTGTCAAGAA | 1079 |
| Db | 301 | AAAAACATGAAATACCCCATACAAAGTCATTTAAATGTACATTTGAAATTTGTCAAGAA | 360 |
| Qy | 1080 | GCAATTTTAAAAACATCAATCTTTAAGATCATATATTAATCTGTTCAGAGAAAAACATTA | 1133 |
| Db | 361 | GCAATTTTAAAAACATCAATCTTTAAGATCATATATTAATCTGTTCAGAGAAAAACATTA | 420 |
| Qy | 1140 | ACGTGTAAACATGTATATPAAAGTTTTCACCTGCACCTTCAAAATTTAGCACACATAAATTA | 1199 |
| Db | 421 | ACGTGTAAACATGTATATPAAAGTTTTCACCTGCACCTTCAAAATTTAGCACACATAAATTA | 480 |
| Qy | 1200 | AAACATCATGGTGCATCTCTGCTTATCATATGTATCATCTGTGTTGTTAAAAATTC | 1259 |
| Db | 481 | AAACATCATGGTGCATCTCTGCTTATCATATGTATCATCTGTGTTGTTAAAAATTC | 540 |
| Qy | 1260 | CAAACTTGAGATTAACAATTCATATTAACAATGCATCCAAAACCTTAATGTCTCT | 1319 |
| Db | 541 | CAAACTTGAGATTAACAATTCATATTAACAATGCATCCAAAACCTTAATGTCTCT | 600 |
| Qy | 1320 | AAATGTGTAAAGTTGTGTGGGAAAAAGTTATCTTCACATATGTTAAGTCATGAT | 1379 |
| Db | 601 | AAATGTGTAAAGTTGTGTGGGAAAAAGTTATCTTCACATATGTTAAGTCATGAT | 660 |
| Qy | 1380 | GATTCATCATGATCACAATATGAGACTTGTGATTTTGTGATGTGTGGGAAATTTGCAAG | 1439 |
| Db | 661 | GATTCATCATGATCACAATATGAGACTTGTGATTTTGTGATGTGTGGGAAATTTGCAAG | 720 |
| Qy | 1440 | AAAAATGATTTAGTTGAAACCTTAATATCTTCATATATGTAAATATTCCTGATGATTTA | 1499 |
| Db | 721 | AAAAATGATTTAGTTGAAACCTTAATATCTTCATATATGTAAATATTCCTGATGATTTA | 780 |
| Qy | 1500 | TTAAAGGAAACCTGAAAGTGAATAAATTTAGAGAACCTATTTAGTCAAGAGATCGAAATTAAT | 1559 |
| Db | 781 | TTAAAGGAAACCTGAAAGTGAATAAATTTAGAGAACCTATTTAGTCAAGAGATCGAAATTAAT | 840 |
| Qy | 1560 | AATTTGATGATTTAGAAACAGGAATTTAAAGTGAAGAGATGAGAAAGATGAAGAA | 1619 |
| Db | 841 | AATTTGATGATTTAGAAACAGGAATTTAAAGTGAAGAGATGAGAAAGATGAAGAA | 900 |
| Qy | 1620 | GATAGCTAGATGAAAAAAGAAATGATTAAGTACAGCTCAATGTCTCAAGATCA | 1679 |
| Db | 901 | GATAGCTAGATGAAAAAAGAAATGATTAAGTACAGCTCAATGTCTCAAGATCA | 960 |
| Qy | 1680 | ATTAATCATTTACTGCTCTTTGGAAAGTTTCAAGAGTGTCTTAACCTTAATTCGAAT | 1739 |
| Db | 961 | ATTAATCATTTACTGCTCTTTGGAAAGTTTCAAGAGTGTCTTAACCTTAATTCGAAT | 1020 |
| Qy | 1740 | AGTGGGAAGAAATCAATTTGCTTAAGATTAATTTGTGATPAAATGTTTTCTAGAGATAT | 1799 |
| Db | 1021 | AGTGGGAAGAAATCAATTTGCTTAAGATTAATTTGTGATPAAATGTTTTCTAGAGATAT | 1080 |
| Qy | 1800 | GATTTACGTGCATTTGAAATGGCATGATPAAATTTCAAGAGATTAAGTATCTTA | 1859 |
| Db | 1081 | GATTTACGTGCATTTGAAATGGCATGATPAAATTTCAAGAGATTAAGTATCTTA | 1140 |

QY 1860 AATAGTATGAAAGAAAGAAAGAACTCCAGAAAGGTGAACCTTGGTTAAAGAAAGCCAGATG 1913
 DB 1141 AATAGTATGAAAGAAAGAAAGAACTCCAGAAAGGTGAACCAATGGTTAAAAAGCCAGATG 1200
 QY 1920 GATTATGCGCAATGAAACATGAGTATTCCTCGATTA 1958
 DB 1201 GATTATGCGCAATGAAACATGAGTATTCCTCGATTA 1239
 RESULT 3
 LOCUS AX489193 1239 bp DNA linear PAT 16-AUG-2002
 DEFINITION Sequence 6493 from Patent WO02053728.
 ACCESSION AX489193
 VERSION AX489193.1 GI:22323205
 KEYWORDS
 SOURCE
 ORGANISM
 Candida albicans
 Candida albicans
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Microsporid; Saccharomycetales; Candida.
 REFERENCE
 1 Roemer, T., Jiang, B., Boone, C., Bussey, H. and Ohlsen, K.L.
 Gene disruption methodologies for drug target discovery
 Patent: WO 02053728-A 6493 11-JUL-2002;
 Elitza Pharmaceuticals, Inc. (US)
 FEATURES
 source
 1..1239
 /organism="Candida albicans"
 /mol_type="unassigned DNA"
 /db_xref="taxon:5476"
 ORIGIN
 Query Match 60.0%; Score 1235.8; DB 6; Length 1239;
 Best Local Similarity 99.8%; Pred. No. 8.2e-187;
 Matches 1237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 720 ATGAGTGAAGTGAACGAAACCAATCGATATCATCTTTAATCTTCTCTTCATCA 779
 DB 1 ATGAGTGAAGTGAACGAAACCAATCGATATCATCTTTAATCTTCTCTTCATCA 60
 QY 780 CGTCCCAAAAGATATTTTGCATATGAAGGCTGTATGAAGCCCTTAATCGACATCA 839
 DB 61 CGTCCCAAAAGATATTTTGCATATGAAGGCTGTATGAAGCCCTTAATCGACATCA 120
 QY 840 TTATTAGAGCAACATTTAAGAACCCACATGATGATGACCGTAAATGACGATGAC 899
 DB 121 TTATTAGAGCAACATTTAAGAACCCACATGATGATGACCGTAAATGACGATGAC 180
 QY 900 GATTGTGATTAAGCATTTTTCAGAAATTCATTTGGAAACACATATTTGATCATTTCC 959
 DB 181 GATTGTGATTAAGCATTTTTCAGAAATTCATTTGGAAACACATATTTGATCATTTCC 240
 QY 960 GAAAAAAGCAATTCATTTGTTGCTAGTGTGTAAAGGGTTAAATTCGCAACACTTG 1019
 DB 241 GAAAAAAGCAATTCATTTGTTGCTAGTGTGTAAAGGGTTAAATTCGCAACACTTG 300
 QY 1020 AAAAGACATGAATACCCCATCAAAAGTCAATTTAAATGTACATTTGAAATTTGCAAGA 1079
 DB 301 AAAAGACATGAATACCCCATCAAAAGTCAATTTAAATGTACATTTGAAATTTGCAAGA 360
 QY 1080 GCATTTTATAAATCATCTTTAAGACATCATATATATCTGTTCAATGAAAAACATTA 1139
 DB 361 GCATTTTATAAATCATCTTTAAGACATCATATATATCTGTTCAATGAAAAACATTA 420
 QY 1140 ACGGTAAACATGATTAATAGTTTTCATCGACCTTCAAAATTAAGCAACATTAATTA 1199
 DB 421 ACGGTAAACATGATTAATAGTTTTCATCGACCTTCAAAATTAAGCAACATTAATTA 480
 QY 1200 AAACATCATGATGATCTCTGCTTATCAATGTGATCATCTGCTGTTGTTTAAAAATTC 1259
 DB 481 AAACATCATGATGATCTCTGCTTATCAATGTGATCATCTGCTGTTGTTTAAAAATTC 540

QY 1260 CAAACTTGCTCAGTATTAATCAATTTCAATATAAACAATCGATCCAAACTTAATCTCT 1319
 DB 541 CAAACTTGCTCAGTATTAATCAATTTCAATATAAACAATCGATCCAAACTTAATCTCT 600
 QY 1320 AAATGTGTAAGGTTGTGTGGAAAAAGTTTATCTTTCACATATGTTAATGCTATG 1379
 DB 601 AAATGTGTAAGGTTGTGTGGAAAAAGTTTATCTTTCACATATGTTAATGCTATG 660
 QY 1380 GATTCTACCATGATCAAAATATGGAATCTGTGATTAATTTGATGATGGGAAATTTGCAAG 1439
 DB 661 GATTCTACCATGATCAAAATATGGAATCTGTGATTAATTTGATGATGGGAAATTTGCAAG 720
 QY 1440 AAAAATGAATTAATGTAACATTAATATCTTCATGATGTAATATCCCTGATGATTA 1499
 DB 721 AAAAATGAATTAATGTAACATTAATATCTTCATGATGTAATATCCCTGATGATTA 780
 QY 1500 TTAAGGAAACCTGAAGTGAAGAAATTTAGAAACCTTTTGAATCAAGATCGAAATTAAT 1559
 DB 781 TTAAGGAAACCTGAAGTGAAGAAATTTAGAAACCTTTTGAATCAAGATCGAAATTAAT 840
 QY 1560 AATTTGCATGAATTAAGAAACAGAAATTAAGAGTGAAGAAAGATGAAGAA 1619
 DB 841 AATTTGCATGAATTAAGAAACAGAAATTAAGAGTGAAGAAAGATGAAGAA 900
 QY 1620 GATAGTCTAGATGAAAAAGAAAGTGAATGTTAGATCAAGCTCAATGTGATCAAAATCA 1679
 DB 901 GATAGTCTAGATGAAAAAGAAAGTGAATGTTAGATCAAGCTCAATGTGATCAAAATCA 960
 QY 1680 ATTAATTCATTTTCTGCTCTTTTGAAGGTCAAGAGTGTCTTCAACTTATCTGAAT 1739
 DB 961 ATTAATTCATTTTCTGCTCTTTTGAAGGTCAAGAGTGTCTTCAACTTATCTGAAT 1020
 QY 1740 AGTGGGAGAAATCAATTTGCTTAAGAAATATTTGATGAATGTTTCTAGAGAAAT 1799
 DB 1021 AGTGGGAGAAATCAATTTGCTTAAGAAATATTTGATGAATGTTTCTAGAGAAAT 1080
 QY 1800 GATTTCAGTGAATTTGAAATGCAATGATGATTAATTAACAAAGATTTGATCTTCTTA 1859
 DB 1081 GATTTCAGTGAATTTGAAATGCAATGATGATTAATTAACAAAGATTTGATCTTCTTA 1140
 QY 1860 AATAGTATGAAAGAAAGAAAGAACTCCAGAAAGGTGAACCAATGGTTAAAAAGCCAGATG 1919
 DB 1141 AATAGTATGAAAGAAAGAAAGAACTCCAGAAAGGTGAACCAATGGTTAAAAAGCCAGATG 1200
 QY 1920 GATTATGCGCAATGAAACATGAGTATTCCTCGATTA 1958
 DB 1201 GATTATGCGCAATGAAACATGAGTATTCCTCGATTA 1239
 RESULT 4
 YSCNOFEAT
 LOCUS 1560 bp DNA linear PLN 27-APR-1993
 DEFINITION Saccharomyces cerevisiae transcription factor IIRA and RNA
 polymerase subunit RPB6 gene.
 ACCESSION M90638.1 GI:172046
 VERSION M90638.1
 KEYWORDS RNA polymerase; transcription factor IIRA.
 SOURCE Saccharomyces cerevisiae (baker's yeast).
 ORGANISM Saccharomyces cerevisiae
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetales; Saccharomycetes.
 REFERENCE
 1 (bases 1 to 1560)
 Moychik, N.A. and Young, R.A.
 Genes encoding transcription factor IIRA and the RNA polymerase
 common subunit RPB6 are divergently transcribed in Saccharomyces
 cerevisiae
 Proc. Natl. Acad. Sci. U.S.A. 89 (9), 3999-4003 (1992)
 JOURNAL MEDLINE
 PUBMED 1570325
 COMMENT
 FEATURES
 source
 1..1560
 /organism="Saccharomyces cerevisiae"

/mol_type="genomic DNA"
/strain="S288C"
/db_xref="taxon:4932"
ORIGIN chromosome 14.

Query Match 10.6%; Score 217.4; DB 8; Length 1560;
Best Local Similarity 59.4%; Pred. No. 4.4e-25;
Matches 409; Conservative 0; Mismatches 271; Indels 9; Gaps 2;

QY 781 GTCCAAAAGATATATTTGACATATGAAAGGGTGTGATTAAGCCCTATATGACCATCAT 840
DB 368 GGGCAAGACATATTTCTGTACTATGATGGCTGTGATTAAGGCAATTTACAGACCTTCAA 427
QY 841 TATTAGGCAACATTTTAAGAACCCACAGTATGATCGACCGTATTAATGTACAGTGACG 900
DB 428 TTTGACTGAACACCAATTAAGCGTACATCAGGGCTTTA--AGAGCATTTTCAGTGTATA 484
QY 901 ATTGTGATAAAGCATTTTTCAGAAAATCACATTTTGAAACACATTTGTATACATTTCCG 960
DB 485 AGTGTGCAAAATCCTTGTATTAAGAGTCACTTAGAGACACTTGTATACGCAATTCG 544
QY 961 AAAAAAACCATTCCATTTGTCAGTGTGTGTAAGGGGTTAAATTCGACCAACTTGA 1020
DB 545 ATACGAACCATTCATTTCTTATTTGTGAAAAGGAGTGAAGACTCGCCAGCACTGA 604
QY 1021 AAGACATGAATCAACCATACAAAGTCATTTTAATGTATCAATTGAAATTTGCAAGAAG 1080
DB 605 AGGACACGAAAGTAAAGCAACCAATCTTTCATTTGTCCAGAGAAAGATGCAACTCC 664
QY 1081 CATTTTATAAACAATCATCTTTTAAGACATCATATATATCTGTCATGAAAAACATTTA 1140
DB 665 GATTTACAAAGATCCACATTTAAGGACATATTTTATCTGTCATTTACATTAACATA 1200
QY 1141 CGGTAAACAATGATTAATAAGTTTTCACCTGACCTTCMAAATTAAGCAACATTAATA 1260
DB 725 CCTGTCCACACTGCATAAAGCTTTCAGAGGCCATATAGGCTTAAGAAATCAATTTCTA 784
QY 1201 AATATCA-----TGGTGAATCTCTGCTTATCAATGTATCAATCTGCTGTTGTTAAA 1254
DB 785 AATATCAAGATCTGTAGAGTGAAGAAATCTTACCAATGTATCTTGTGCTGTTGCAAG 844
QY 1255 AATTCACAACTTGATGATTTACATTTTACATTAATAACAATCCAAACCTTAAT 1314
DB 845 AGTTTGGATATGTGTCAATTTGCAATTCGATTTAAAAATGATCATCTCAAAATTAATA 904
QY 1315 GTCTTAATAGTGTAAAGGTTGTGTTGGAAAAAGGTTATCTTCACATATGTAAAGTC 1374
DB 905 GTCCATATTTGTAGCAAAACCATGTGTGGGGGAAATGTTTACAAATGACATGATTAATC 964
QY 1375 ATGATGATTTCAACATGATCAAAATATGACATTTGTGATTTGTGATGTGGGAAATTTG 1434
DB 965 ATGACACTCACTAGTAAACCAAAATTTGAAAGTGTATATGTCTCGATATGTCTTTT 1024
QY 1435 CAAGAAAAATGATTAAGTGAACATTAAT 1463
DB 1025 CTAGAAAACATGATCTTCTACGCAATTAAT 1053

RESULT 5
YSCFIIIA 1739 bp DNA linear PLN 13-SEP-1996
LOCUS Saccharomyces cerevisiae transcription factor IIfa (TFIIfa) gene,
DEFINITION complete cds.
ACCESSION M80611.1 GI:172902
VERSION M80611.1
KEYWORDS transcription factor IIfa; zinc-finger protein; zinc-finger
SOURCE Saccharomyces cerevisiae (baker's yeast)
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE 1 (bases 1 to 1739)
AUTHORS Archambault,J., Milne,C.A., Schappert,K.T., Baum,B., Friesen,J.D.

TITLE and Segall,J.D.
The deduced sequence of the transcription factor TFIIfa from
Saccharomyces cerevisiae reveals extensive divergence from Xenopus
TFIIfa
JOURNAL U. Biol. Chem. 267 (5), 3282-3288 (1992)
MEDLINE 92147684
PubMed 1737784
FEATURES
source

Location/Qualifiers
1..1739
/organism="Saccharomyces cerevisiae"
/mol_type="genomic DNA"
/strain="M303-1A"
/db_xref="taxon:4932"
/chromosome="XVI"
234..1523
/gene="TFIIfa"
234..1523
/gene="TFIIfa"
/note="zinc finger containing protein"
/codon_start=1
/product="transcription factor IIfa"
/protein_id="AA08014.1"
/db_xref="GI:172902"
/translation="MGEEVLANNEGMPLEIKQETIPISRESESEINSITSTSSSN
RPTVPCDYDCDKAFTRPSILTEHQLSVQGLRAEQCKAKSPVKSHLEHLYTH
SDTPKQCYCGKGVTRROOLKREHVTHTKSFICPEGCNLRYPKHQOLRAHLYTH
HKLTPCHNKSFORPYRLRNHISKHDPEVNPYOCCTPAGCKEPRIMSOLOSHIKND
HPELKPDISKPCVGENGIOMNMIHDDSLVYTNMKCHICPDMSFSKRELLTHYGI
HPEEDIPLEIKRISDIQOLVDHGTVQLGSKNSODEKISNLRKRLKENNV
EFLQNVLDLKRLESENGJNLNLTAVTGKRYRCFYNNCSRTFTKRYEKHIDKHVH
ELKTLLOREKENKTLVDONHKEPFIIOKETOSAGDK"

ORIGIN

Query Match 10.6%; Score 217.4; DB 8; Length 1739;
Best Local Similarity 59.4%; Pred. No. 4.4e-25;
Matches 409; Conservative 0; Mismatches 271; Indels 9; Gaps 2;

QY 781 GTCCAAAAGATATATTTGACATATGAAAGGGTGTGATTAAGCCCTATATGACCATCAT 840
DB 367 GGGCAAGACATATTTCTGTACTATGATGGCTGTGATTAAGGCAATTTACAGACCTTCAA 426
QY 841 TATTAGGCAACATTTTAAGAACCCACAGTATGATGACCGTATTAATGTACAGTGACG 900
DB 427 TTTGACTGAACACCAATTAAGGCTATCAGGGTTTA--AGAGCATTTTCAGTGTATA 483
QY 901 ATTGTGATAAAGATTTTTCAGAAATCATCTTTGAAACACATATTTGATCAATTTCCG 960
DB 484 AGTGTCAAAATCTTCTGTTAAAAAGAGTCACTTGAAGACACATTTGATACGCAATTCG 543
QY 961 AAAAAAACCATTCCATTTGTCAGTGTGTGTAAGGGGTTAATCTCGACCAACATTTGA 1020
DB 544 ATAGCAAGACATTTCCATATTTCTTATTTGTGAAAAAGAGTACGACTCGCCAGCACTGA 603
QY 1021 AAGACATGAATCAACCCATACAAAGTCATTTAAATGTACATTTGAAATTTGTCAAGAAG 1080
DB 604 AGCGACACCAAGTAAAGCATACCAAAATCTTCAATTTGTCCAGAGAAAGATGCAACCTCC 663
QY 1081 CATTTTATAAACAATCATCTTTTAAGACATCATATATATCTGTCATGAAAAACATTTA 1140
DB 664 GATTTCTACAGCATCCACATTTAAGGGCACATATTTTATCTGTCATTTACATTAACATA 1200
QY 1141 CGGTAAACAATGATTAATAAGTTTTCACGACCTTCAAAATTAAGCAACATTAATA 1260
DB 724 CCTGTCCACACTGCATAAAGACCTTTCAGAGGCCATATAGGCTTAAGAAATTCATTTCTA 783
QY 1201 AATATCA-----TGGTGAATCTCTGCTTATCAATGTATGATCATCTGCTGTTTAAA 1254
DB 784 AATATCAAGATCTGTAGAGTGAAGAAATCTTACCAATGATCTTTGTGCTGTTGTGCAAG 843
QY 1255 AATTCACAACTTGATGATTTACATTTTCAATTAATAACAATGATCCAAACCTTAAT 1314
DB 844 AGTTTGGATATGTATCAATTTGCAATGACATTAATAAATGATATCTCTTAATTAAT 903

| | | | | |
|-------------------|--|----------|--|--------|
| Oy | | 1315 | GTCTAAATGCGTAACGGTTGGTGGGAAAAAGTTTTATCTTCACAATATGTTAATGC | 1274 |
| Dd | | 904 | GTCCTATTGTTGAAGCAACCAGTGTTGGGGAAAATGGTTTACAAATCAGCATGATTATTC | 963 |
| Oy | | 1375 | ATGATGATTTCACCCAGTAGTCATAAATATATGACTTGATTAATTGATGATGGGAAAATTG | 1434 |
| Dd | | 964 | ATGACGACTACCTAGTAAACCAAATAATTGGAAGTGTCAATAATATGCTCATATATGCTTTT | 1023 |
| Oy | | 1435 | CANAAGAAAATGCAATTAGTTGAACATTAT | 1463 |
| Dd | | 1024 | CTAGA AAAACATGATCTTCTCACGATTAT | 1052 |
| RESULT 6 | | | | |
| YSCP9677/c | | | | |
| LOCUS | YSCP9677 | 37497 bp | DNA | linear |
| DEFINITION | Saccharomyces cerevisiae chromosome XVI cosmid 9677. | | | |
| ACCESSION | U25841 U00094 | | | |
| VERSION | U25841.1 | | | |
| KEYWORDS | GI:786295 | | | |
| SOURCE | . | | | |
| ORGANISM | Saccharomyces cerevisiae (baker's yeast) | | | |
| | Saccharomycetes cerevisiae | | | |
| | Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; | | | |
| | Saccharomycetales; Saccharomycetaceae; Saccharomycos. | | | |
| REFERENCE | 1 (bases 1 to 37497) | | | |
| AUTHORS | Johnston,M., Andrews,S., Brinkman,R., Cooper,J., Ding,H., Du,Z., Favelllo,A., Fulton,L., Gattung,S., Greco,T., Kirsten,J., Kucaba,T., Hallsworth,K., Hawkins,J., Hillier,L., Jier,M., Johnson,D., Johnston,L., Langston,Y., Latreille,P., Le,T., Mardis,E., Meneses,S., Miller,N., Nhan,M., Pauley,A., Peluso,D., Rifken,L., Riles,L., Tatch,A., Trevasakis,B., Vignati,D., Wilcox,L., Wohlman,P., Vaudin,M., Wilson,R. and Waterston,R. | | | |
| TITLE | The sequence of a portion of the right arm of Saccharomyces cerevisiae chromosome XVI unpublished (1995) | | | |
| JOURNAL REFERENCE | 2 (bases 1 to 37497) Miller,N. | | | |
| AUTHORS | The sequence of S. cerevisiae cosmid 9677 unpublished (1995) | | | |
| JOURNAL REFERENCE | 3 (bases 1 to 37497) Waterston,R. | | | |
| AUTHORS | Direct Submission Submitted (27-APR-1995) Robert Waterston | | | |
| TITLE | Submitted by: | | | |
| JOURNAL COMMENT | Genome Sequencing Center Department of Genetics, Washington University, St. Louis, MO 63110, USA e-mail: mj@sequencer.wustl.edu | | | |
| FEATURES | source | | | |
| | NEIGHBORING COSMID INFORMATION: This sequence includes nucleotides 1-37497 of cosmid 9677. The cosmid on the left is 9705. Location/Qualifiers 1..37497 organism="Saccharomyces cerevisiae" mol_type="genomic DNA" strain="S288C (AB972)" db_xref="taxon:4932" map="XVI" note="Baker's Yeast" complement(2366..3655) gene="PFC2" complement(2366..3655) name="PFC2" gene="PFC2" note="P9677.g" | | | |
| | /codon_start=1 /product="transcription factor TFIIIA (PIR accession number S20050)" /protein_id="BAB64615.1" /db_xref="GI:786305" /translation="MGGEVLANNEGKPLAEIKQETPIPRSSSSBSINSITGRSSSNRPKYTFICYDGDKAFTFPOSLIRHEVTSTKSFI CEDEBCNIRFYKHDPILRAHLSVHI SDTRPKSYSGKGVTTPQQDLRHVVTHTKSFICPEBCCNIRFYKHDPILRAHLSVHI | | | |

[illegible]

Query Match 10.6%; Score 217.4; DB 8; Length 37497;
Best Local Similarity 59.4%; Pred. No. 2,66-25;
Matches 409; Conservative 0; Mismatches 271; Indels 9; Gaps 2;

| Accession | Version | Keywords | Source | Organism | Reference | Authors | Title | Journal | Medline | Pubmed | Comment |
|-------------|--|---|--------|----------|-----------------|---------|-------|---------|---------|--------|---------|
| QY | 781 | GTCCCAAAAAGTATATTTCGACATATGAAAGGGTGTGATAAAGCCCTATATGACCATCAT | 840 | | | | | | | | |
| Db | 3322 | GGCCAAAGACATATTTCTGTGACTATGATGCGTGTGATTAAGGACATTTTACAAGACCTTCAA | 3463 | | | | | | | | |
| QY | 841 | TATTAGAGCAACATTTTAAAGAACCCACAGTAATATATGACCGGTATTAATGTACAGTGGACG | 9003 | | | | | | | | |
| Db | 3462 | TTTTGACTGACACACCAATTAAAGGTACATATAGGGTTTAA--AGAGCATTTTCAGTGTGATTA | 3404 | | | | | | | | |
| QY | 901 | ATTGTGATTAAGACATTTTTCAGAAAATTCATTTTGGAAAACACATATTTGTATCATCTCCG | 9666 | | | | | | | | |
| Db | 3405 | AGGTGGCAAAATCCTTCGTTAAAAAGAGTCACTTAAGAGACACCTTGTAACGATTTCTG | 3334 | | | | | | | | |
| QY | 961 | AAAAAAAACCATTTCCATTTGTCAGTGTGTGTGATAAAGGGGTTAAATTCGACCAACACTTGA | 1102 | | | | | | | | |
| Db | 3345 | ATACGAACCATTTCCATATGTTCTTATTTGTGAAAAAGAGTGAACGACTCGGACGACACTGA | 328 | | | | | | | | |
| QY | 1021 | AAAGACATGAATCACCCCATPACAAAGTCAATTTTAAAGTACATTTGAAAATTTGTCAAGAA | 108 | | | | | | | | |
| Db | 3285 | AGGACACGACAGTAAAGCATATCCAAATCTTTTCATTTGTCCAGAGAAAGATGACACTCC | 322 | | | | | | | | |
| QY | 1081 | CATTTTATAAACATCAATCTTTTAAAGACATCATATATATCTGTTCATGAAAAACAATTAA | 114 | | | | | | | | |
| Db | 3225 | GATTTTACAAGCATCCACAAATTAAGGGCACATATTTTACTGTTCATTTACATTAACCTAA | 316 | | | | | | | | |
| QY | 1141 | CGGTGAACAAATGTATTAAGAATTTTCACTGCACCTTCAAAATTAAGCAACAACTAAATTA | 1200 | | | | | | | | |
| Db | 3165 | CCTGTCCACATCGCAATTAATAAGCTTTCAAGAGCCCATATATGGCTTAAGAAATCACATTTCTA | 3100 | | | | | | | | |
| QY | 1201 | AACATCA-----TGGTGGATCTCGCTTATTCATATGATCATATCCTGGTGTTTTAA | 125 | | | | | | | | |
| Db | 3105 | AACATCAGATCTCTAGGTGAGAAATCCTTACCAATGTATCTTTGCTGGTGTTCGAAG | 3044 | | | | | | | | |
| QY | 1255 | ATTTCAAACTTGGTCACTTATTAACAATTCATATTAACCACTGATCCAAACTTAAAT | 1314 | | | | | | | | |
| Db | 3045 | AGTTTCGATATGTCTACCAATTCGACATATATAAAAAATATATCTTAATTTAAAT | 2988 | | | | | | | | |
| QY | 1315 | GTCCTAATGTGTAAAGTGTGTGTGGAAAAAGTTTATCTTCAATATGTTAAGTC | 1377 | | | | | | | | |
| Db | 2985 | GTCTTAATTTGAGCAACCATGTGTGTGGGAAAAAGTTTACAAATGACATGATTAATTC | 2922 | | | | | | | | |
| QY | 1375 | ATGATGATTTCAACATGATCAAAAATATATGAACTGTGATTTATGTGATGTGGGAAATTTG | 1433 | | | | | | | | |
| Db | 2925 | ATGAGAGCTCACTATATACCAAAAATTTGAAAGTGTCAATATATGTCTGATATGTCTTTT | 2866 | | | | | | | | |
| QY | 1435 | CAAGAAAAATGCAATTAAGTTGAACATTAAT | 1463 | | | | | | | | |
| Db | 2865 | CTAGAAAAACATGATCTTCTCACGATTAAT | 2837 | | | | | | | | |
| RESULT 7 | | | | | | | | | | | |
| YSCRPO26A/c | | | | | | | | | | | |
| LOCUS | YSCRPO26A | 2133 bp | DNA | linear | PLN 27-APR-1993 | | | | | | |
| DEFINITION | S.cerevisiae RNA polymerase II sixth subunit (Rpo26) gene, complete | | | | | | | | | | |
| ACCESSION | M33924 | | | | | | | | | | |
| VERSION | M33924.1 | GI:172452 | | | | | | | | | |
| KEYWORDS | RNA polymerase II; c-myc proto-oncogene; transmembrane protein; tyrosine kinase. | | | | | | | | | | |
| SOURCE | Saccharomyces cerevisiae (baker's yeast) | | | | | | | | | | |
| ORGANISM | Saccharomyces cerevisiae | | | | | | | | | | |
| REFERENCE | Bukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces. | | | | | | | | | | |
| AUTHORS | Archambault,J., Schappert,K.T. and Friesen,J.D. | | | | | | | | | | |
| TITLE | A suppressor of an RNA polymerase II mutation of Saccharomyces cerevisiae encodes a subunit common to RNA polymerases I, II, and III | | | | | | | | | | |
| JOURNAL | Mol. Cell. Biol. 10 (12), 6123-6131 (1990) | | | | | | | | | | |
| MEDLINE | 91064718 | | | | | | | | | | |
| PUBMED | 2247052 | | | | | | | | | | |
| COMMENT | Original source text: S.cerevisiae single-copy plasmid DNA, clone 2. | | | | | | | | | | |

Query 957 TCCGAAAAAACCATTCATTTGTCAGTGTGTTAAAGGGTTAATTCGCAACAC 1016
 Db 143 TCTATTCGAGGCAATTCATTTGTTCTTATGTGCAAGGGGGTGACAACTCGACAGCAA 202
 QY 1017 TTGAAAAGACATGAAATCACCACATCAAAAGTCATTAAATGTAATTTGAAATGTCOA 1076
 Db 203 TTGAAGGGCCAGAAAGTTACACATACCAAACTTTTGTTCGGAAGATGATGAT 262
 QY 1077 GAACATTTTAAACATCAATCTTTAAACATATATATATCTGTCATGAAAAACA 1136
 Db 263 CTCGATTTTAAAGCAACCTCAATTAAGGGCTCATATCTGTCACCTAGACAA 322
 QY 1137 TTAAGTGTAAACATGTAATTAAGTTTCACTGACCTTCAAAATTAAGACAACTAAA 1196
 Db 323 TTAACCTGTCCACATGTAATTAATTTTCAAGACCTTACAGGCTTAAAGAAATCATATC 382
 QY 1197 TTAACATCATGTGT-----GGATCTCTGCTTATCATGATCATCTGCTGTGTTT 1250
 Db 383 TCTAAACATCATGATCTGTAAGTGTAAATCCGATTAATGATCTTCTGCTGTGTT 442
 QY 1251 AAAAATTTCCAACTGTGTGATGATTAACAATTTCAATTAACAACTGCATCCAAAATT 1310
 Db 443 ACAGAGTACCGTAAATGTGTACAGTTGCATCATATCAAAATGACATCTTAATTG 502
 QY 1311 AATGTCTTAATAGTGTAAAGTGTGTTGGAAAAAGTTTATCTTCAATATGTTA 1370
 Db 503 AATGCCCCATTTGCAACAGCTTGTGTAGGGGAAATGTCTCAAAATGCATATGATC 562
 QY 1371 AGTCATGATGATTCATACATGATCAAAATATGACATGTGATATGATGATGGGAAA 1430
 Db 563 ATTCATGATGATCTTGTGTAAGTAAAGTGAATGCAATGCAATTTGTTCTAATACATCT 622
 QY 1431 TTGCAAAAGAAAAATTAATTAAGTTGAACATTAATATATCTTCAT 1475
 Db 623 TTCTTGAAGAACAGACCTCTTACTACTACACAGATGACCAT 667

RESULT 9
 AC138126/c 169063 bp DNA linear PRI 17-DEC-2002
 LOCUS AC138126
 DEFINITION Homo sapiens chromosome 19 clone RP11-274A19, complete sequence.
 AC138126.1 GI:27151357
 VERSION
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 169063)
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
 TITLE Direct Submission
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 169063)
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
 TITLE Direct Submission
 JOURNAL Submitted (17-DEC-2002) DOE Joint Genome Institute, 2800 Mitchell
 Drive, Walnut Creek, CA 94598, USA
 COMMENT Draft Sequence Produced by DOE Joint Genome Institute
 www.jgi.doe.gov

FEATURES
 source
 Location/Qualifiers
 1..169063
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="19"
 /clone="RP11-274A19"

ORIGIN

Query Match 4.9%; Score 101; DB 9; Length 169063;
 Best Local Similarity 50.8%; Pred. No. 6,1e-07;
 Matches 298; Conservative 0; Mismatches 280; Indels 9; Gaps 2;

QY 726 GAAAGTACGAAACCAATGATATCATCTTTATATCTTCTTCTATCAGCGCCC 785
 Db 80134 GAATGTGGCAAGCTTTAAACAGCTCTCATCTTACTACATAGAAATTCATCT 80075
 QY 786 AAAAGATATTTGACATATGAAGGTGTGATTAAGCTTAATGACATCTTATTA 845
 Db 80074 GAGAGAAATCTTCAAAATGTGAAGATGTGGCAAAAGCTTTATATCATCTTCAAACTT 80015
 QY 846 GAGCAACATTTAAGAACCCACAGTAATGATGACCGGTAATTAAGTACAGTGGATTT 905
 Db 80014 ACTGAACATTAAGAAATTCATCTGAGAGAAACCTTACATGT-----GAAGATGT 79961
 QY 906 GATTAAGCATTTTTCAGAAAATGACATTTGGAACACATATGATATCATCTTCAAAA 965
 Db 79960 GGCAAAGCTTTTAAACATTTCTTCAATCTTCAATCTTCAATAGTATCTGAGAG 79901
 QY 966 AAAACCATTCATTTGTCAGTGTGTGTAAGGGGTTAATTTGACACACTTGAAGA 1025
 Db 79900 AAACCTTACAAATGTAAGAAATGTGTAAAGCTTTAAACAGTCTTCAACCTTACAG 79841
 QY 1026 CATGAATCACCCATCAAAAGTCATTAATGATGATGAAATTTGCAAGACATTT 1085
 Db 79840 CATTAAGAAATTCATCTGAGAGAAACCTTCAATGTAAGAAATGTGCAAAAGCTTTT 79781
 QY 1086 TATTAACATCAATCTTTAA--GACATCATATATATCTGTCATGAAAAACATTAAG 1142
 Db 79780 AACCAAGTCTTCAACCTTACTGACATTAAGAAATTCATCTGATGAGAAACCTTCAAA 79721
 QY 1143 TGTAAACATGTAATTAAGTTTTCAGTGTGCACTTCAAAATTAAGACACATTAATTA 1202
 Db 79720 CTTAAAGATGTAAAGATGATTTGAAAACCTTCAAAAGTTTCTTAACATTAAGAAAT 79661
 QY 1203 CATCATGTGATCTCTGCTTATCATATGATCATCCGCTGTTTAAATTTCCAA 1262
 Db 79660 TATGCTGTGTAAGAAATCTTGAAGATGTGAAGATGAACAAACCTTTAAAGGTGTAC 79601
 QY 1263 ACTTGTCATTAATTAATTTCAATTAACAACTGCATCCAAACT 1309
 Db 79600 ACTTGATGTGATGAATTAATTTCACTTAATAAAAAAACCTTCAAGT 79554

RESULT 10
 AC073544/c 177299 bp DNA linear PRI 19-JUN-2002
 LOCUS AC073544
 DEFINITION Homo sapiens chromosome 19 clone RP11-359H18, complete sequence.
 AC073544.4 GI:21465367
 VERSION
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 177299)
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
 TITLE Direct Submission
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 177299)
 AUTHORS DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (22-JUN-2000) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT On Jun 19, 2002 this sequence version replaced gi:13699752.
 Draft Sequence Produced by DOE Joint Genome Institute
 www.jgi.doe.gov

Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu

Quality: Phrap Quality >=40 99.9% of Sequence;
Estimated Total Number of Errors is 0.2.
NOTE: BACTERIAL TRANSPOSONS excised at 14620 and 98099.

FEATURES

source
1. 177299
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="19"
/clone="RP11-359H18"

ORIGIN

Query Match 4.9%; Score 101; DB 9; Length 177299;
Best Local Similarity 50.8%; Pred. No. 6.1e-07;
Matches 298; Conservative 0; Mismatches 280; Indels 9; Gaps 2;
QY 726 GAAAGTGACGAAACCAATGATATCATCTTTAATCTCTTCTCTTCATCACTCC 785
DB 163740 GAATGGGCAAGCTTTTAACCACTCCTCATCTTACTACACATAGAAATTCATCT 163681
QY 786 AAAAGTATATTTGGCATATGAAAGGTGATAAAGCCTATATGACCATCATTTA 845
DB 163680 GGAGAGAAATCCTACAAATGTGAAGATGTGGCAAGCTTTCTATGATCTCAAACT 163621
QY 846 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATTAATGTACAGTGAAGATTGT 905
DB 163620 ACTGAACATAGAAATTCATCTCGAGAGAAACCTACACATGT-----GAAGATGT 163567
QY 906 GATTAAGCATTTTTCAGAAATACATTTGGAAACACATTTGTATCACTCCGAAAA 965
DB 163566 GGCAGAGCCTTTAACCATCTCCTACATCTGTCTACACATAGATTAATTCATCTGAGAG 163507
QY 966 AAACCATTCATTTGTGAGTGTGTGTAAGGGTTAATCTCGACAAACCTTGAAAAA 1025
DB 163506 AAACCTTACCAATGTGAAGATGTGTAAAGCTTTAAACAGTCTCAACCTTACAGA 163447
QY 1026 CATGAATCAACCCATACAAAGTCATTTAAATGTACATTTGAAATTTGTCAAGAACTTT 1085
DB 163446 CATTAAGAAATTCATCTCGAGAGAAACCTACCAATGTGAAGAAATGTGGCAAGCTTTT 163387
QY 1086 TATTAACATCAATCTTTA--GACATCATATATATCTGTCTCAAGAAAAACATTAACG 1142
DB 163386 AACAGCTCTCAACCTTACTGACATAGAAATTCATCTGTGTGAGAACTCTACAAA 163327
QY 1143 TGTAAACAATGTATTAAGTTTTCATCTGACCTTCAAAATTAAGCAACATTAATTA 1202
DB 163326 CTTAAAGATGTACAGTATTTTGAAGAACTTCAAAAGTTTCTAAACATTAAGAAAT 163267
QY 1203 CATCATGGTGAATCTCTGCTTATCAATGTGATCATCTGTGTGTTTAAATTTCCAA 1262
DB 163266 TATGCTGTGAGAAATCTTGAAGAAATGTGAAGATGTAAACAAACCTTTAAAGTTGTAC 163207
QY 1263 ACTTGTCATGATTTCAATTTCAATTAAGAAACATCGATCCAAACT 1309
DB 163206 ACTGATGTGACATTAATTCATCTAATAAAAAAAACCTTAAGCT 163160

RESULT 11

AX714042 2132 bp DNA linear PAT 15-APR-2003
LOCUS AX714042
DEFINITION Sequence 726 from Patent EP1293569.
ACCESSION AX714042
VERSION AX714042.1 GI:29888970
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

1 Isega,T., Sugiyama,T., Otsubki,T., Wakamatsu,A., Sato,H., Ishii,S., Yamamoto,D.I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R.,

Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Nagahari,K. and Masuko,Y.
Patent: EP 1293569-A 726 19-MAR-2003;
JOURNAL Helix Research Institute (JP) ; Research Association for Biotechnology (JP)

FEATURES

source
1. 2132
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 4.8%; Score 99.4; DB 6; Length 2132;
Best Local Similarity 50.6%; Pred. No. 2.3e-06;
Matches 297; Conservative 0; Mismatches 281; Indels 9; Gaps 2;
QY 726 GAAAGTGACGAAACCAATGATATCATCTTTAATCTCTTCTCTTCATCACTCC 785
DB 1384 GAATGGGCAAGCTTTTAACCACTCCTCATCTTACTACACATAGAAATTCATCT 1443
QY 786 AAAAGTATATTTGGCATATGAAAGGTGATAAAGCCTATATGACCATCATTTA 845
DB 1444 GGAGAGAAATCCTACAAATGTGAAGATGTGGCAAGCTTTCTATGATCTCAAACT 1503
QY 846 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATTAATGTACAGTGAAGATTGT 905
DB 1504 ACTGAACATAGAAATTCATCTCGAGAGAAACCTACACATGT-----GAAGATGT 1557
QY 906 GATTAAGCATTTTTCAGAAATACATTTGGAAACACATTTGTATCACTCCGAAAA 965
DB 1558 GGCAGAGCCTTTAACCATCTCCTACATCTGTCTACACATAGATTAATTCATCTGAGAG 1617
QY 966 AAACCATTCATTTGTGAGTGTGTGTAAGGGTTAATCTCGACAAACCTTGAAAAA 1025
DB 1618 AAACCTTACCAATGTGAAGATGTGTAAAGCTTTTAAACAGTCTCAACCTTACAGA 1677
QY 1026 CATGAATCAACCCATACAAAGTCATTTAAATGTACATTTGAAATTTGTCAAGAACTTT 1085
DB 1678 CATTAAGAAATTCATCTCGAGAGAAACCTACCAATGTGAAGAAATGTGGCAAGCTTTT 1737
QY 1086 TATTAACATCAATCTTTA--GACATCATATATATCTGTCTCAAGAAAAACATTAACG 1142
DB 1738 AACAGCTCTCAACCTTACTGACATAGAAATTCATCTGTGTGAGAACTCTACAAA 1797
QY 1143 TGTAAACAATGTATTAAGTTTTCATCTGACCTTCAAAATTAAGCAACATTAATTA 1202
DB 1798 CTTAAAGATGTACAGTATTTTGAAGAACTTCAAAAGTTTCTAAACATTAAGAAAT 1857
QY 1203 CATCATGGTGAATCTCTGCTTATCAATGTGATCATCTGTGTGTTTAAATTTCCAA 1262
DB 1858 TATGCTGTGAGAAATCTTGAAGAAATGTGAAGATGTAAACAAACCTTTAAAGTTGTAC 1917
QY 1263 ACTTGTCATGATTTCAATTTCAATTAAGAAACATCGATCCAAACT 1309
DB 1918 ACTGATGTGACATTAATTCATCTAATAAAAAAAACCTTAAGCT 1964

RESULT 12

AK056088 2132 bp mRNA linear PRI 01-AUG-2002
LOCUS AK056088
DEFINITION Homo sapiens cDNA FLJ31526 f1s, clone NT2R12000341, moderately similar to ZINC FINGER PROTEIN 43.
ACCESSION AK056088
VERSION AK056088.1 GI:16551397
KEYWORDS oligo capping, f1s (full insert sequence).
SOURCE
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

1 Ninomiya,K., Wagatsuma,M., Kanda,K., Kondo,H., Yokoi,T., Kodaira,H., Furuya,T., Takahashi,M., Kikkawa,E., Omura,Y., Abe,K.,

Kamihara, K., Katsuta, N., Sato, K., Tanikawa, M., Yamazaki, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshihara, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Masuho, Y., Nagai, K. and Isogai, T.
NEBO human cDNA sequencing project
Unpublished
2 (bases 1 to 2132)
Isogai, T., Otsuki, T. and Sugiyama, T.
Direct Submission
Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986) NEBO Human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'-& 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.

FEATURES

Location/Qualifiers

1..2132
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="NT2R12000341"
/cell_line="NT2"
/cell_type="keratocarcinoma"
/clone_lib="NT2R12"
/note="cloning vector: pME18SFL3-mRNA from NT2 neuronal precursor cells treated 2-weeks mitotic inhibitor after 5-weeks retinoic acid (RA) induction..majorly NT2 neuron"
148..1878
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB71090.1"
/db_xref="GI:16551398"
/translation="MVAPPVISCSPHAFHSPSPONIKDSFOKVTTPRYKCHENLOT
SKSVDECKYQKNGNLNOCPLPTQSKIFQCDKYMKI PFKFSNLNGHKYRHRKRPK
YKFGKSCFISNLTQHKICTRVNPKYKEDGKAFNSISITTKRKHITGKSYICE
EGKAKNOFTNLTTHKITITRDKLYKRECSAFNSISHTHTIITHTGNEPKKEEC
DKAFNOSTLTTHKITITHTKINEKNEKCEKAFNOSHLTTHKITHTGKPYCECEGK
AFNOSHLTTHKITHTGKPYCECEGKAFNOSHLTTHKITHTGKPYCECEGK
FNSHLTTHKITHTGKPYCECEGKAFNOSHLTTHKITHTGKPYCECEGKAFNOSHL
FNSHLTTHKITHTGKPYCECEGKAFNOSHLTTHKITHTGKPYCECEGKAFNOSHL
KLTEHKIHTGKPYCECEGKAFNOSHLTTHKITHTGKPYCECEGKAFNOSHL
TTHKITHTGKPYCECEGKAFNOSHLTTHKITHTGKPYCECEGKAFNOSHL
HKRYNAGEKS"

CDS

Query Match 4.8%; Score 99.4; DB 9; Length 2132;
Best Local Similarity 50.6%; Pred. No. 2.3e-06;
Matches 297; Conservative 0; Mismatches 281; Indels 9; Gaps 2;

ORIGIN
Query Match 4.8%; Score 99.4; DB 9; Length 2132;
Best Local Similarity 50.6%; Pred. No. 2.3e-06;
Matches 297; Conservative 0; Mismatches 281; Indels 9; Gaps 2;
Db 726 GAAAGTACGAAACCAATGATATCATCTTAAATATCTTCTTCTCATCAGTCCC 785
1384 GAATGTGGCAAGCTTTTAAACAGTCCCAATCCTTACTACACATAGAGATTACT 1443
Qy 786 AAAAGTATTTTGCATATGAAAGGTGTGATAAGCTTATATGACCATCATATTA 845
Db 1444 GAGAGAAATCTCTAATAATGTGAAAGTGTGCAAGCTTCTATGATCCCAAACTT 1503
Qy 846 GAGCAACATTTAAGAACCCACAGTATGATGACGATTAATATGACAGTGAACATTGT 905
Db 1504 ACTGAACATTAAGAAATCTATCTGAGAGAAACCTCTACATGT-----GAAATAATGT 1557
Qy 906 GATAAGCATTTTTCAGAAATACATTTGAAAACACATATTTGATCATCTCCGAAAAA 965
Db 1558 GGCAGAACCTTTTAAACCACTCTCAACCTTGTCTACACATTAAGTATTCATCTGAGAG 1617

Qy 966 AACCACTTCATTTGTCAGTGTGTGTAAGGGGTATTCGACACACTGTGAAGA 1025
Db 1618 AACCACTTCATTTGTCAGTGTGTGTAAGGGGTATTCGACACACTGTGAAGA 1677
Qy 1026 CATTAATTCACCAATACCAATGATTTAAATGTCATTTGAAAAATGTCAGAAACATT 1085
Db 1678 CATTAATTCACCAATACCAATGATTTAAATGTCATTTGAAAAATGTCAGAAACATT 1737
Qy 1086 TATTAACATCAATCTTTA---GACATCATATATATCTGTTCATGAAAAACATTACG 1142
Db 1738 AACCACTTCATTTGTCAGTGTGTGTAAGGGGTATTCGACACACTGTGAAGA 1797
Qy 1143 TGTAAACATGATTAATAAGTTTTCATCTGACCTTCATTAAGCAACATTAATAA 1202
Db 1798 CTTAAAGATGTAACAGTGTGTTGAAAAACATCTTCAAGTTTCTTAAACATTAAGAAT 1857
Qy 1203 CATCATGTGATCTCTGCTTTCATCAATGTCATCTGTTTAAATTTCCAA 1262
Db 1858 TATGCTGTGAGAAATCTTGAATATGTAAGAAATGTAACAAACCTTTAAAGTTGTAC 1917
Qy 1263 ACTTGTCATATTAATTAATTCATTAATAACACTGCATCCAAACT 1309
Db 1918 ACTTGTCATATTAATTAATTCATTAATAACACTGCATCCAAACT 1964

RESULT 13
AC099500/c 138627 bp DNA linear PRI 29-MAY-2002
LOCUS Homo sapiens chromosome 19 clone RP11-209J6, complete sequence.
DEFINITION AC099500
AC099500.2 GI:21240690
VERSION HTG.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 138627)
DOE Joint Genome Institute and Stanford Human Genome Center.
Unpublished
JOURNAL 2 (bases 1 to 138627)
DOE Joint Genome Institute.
TITLE Direct Submission
AUTHORS Submitted (15-NOV-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
JOURNAL 3 (bases 1 to 138627)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
AUTHORS Submitted (29-MAY-2002) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
JOURNAL On May 29, 2002 this sequence version replaced gi:16930916.
COMMENT Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.1.
NOTE: This insert is not the entire sequence of the clone (entire
sequence is 153.8kb). It is clipped at the overlap with AC008981.
The number of bases overlapped is 27192.

FEATURES

Location/Qualifiers

1..138627
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="19"
/clone="RP11-209J6"

ORIGIN

Query Match 4.8%; Score 99.4; DB 9; Length 138627;
Best Local Similarity 50.6%; Pred. No. 1.1e-06;
Matches 297; Conservative 0; Mismatches 281; Indels 9; Gaps 2;

```

QY 726 GAAAGTACGAAACCAATCGATATCATCTTTAATATCTTCTTCTTCATCAGCTCC 785
DB 48699 GAATGTGGCAAAAGCTTTTAAACGATCCTCAATCTTATCAACATTAAGAAATTCATCT 48640
QY 786 AAAAATTAATTTTGCACATATGAAGGGTGTGATAAGCCCTATATGACCATCATTTA 845
DB 48639 GAGAGAAATCCTACAAATGTGAAGATGTGGCAAGCTTCTCATGATCTCTCAAACTT 48580
QY 846 GAGCAACATTTTAAAGAACCAAGTAAATGATGACCGTATTAATGTAAGTGAAGCATTTGT 905
DB 48579 ACTGAAATTAAGAAATTCATCTGAGAGAAACCTACACATGT-----GAAAGATGT 48526
QY 906 GATTAAGCATTTTTCAGAAAAATGACATTTTGAAGAACATATTTGATACATTCGCAAAA 965
DB 48525 GGCAGAGCCCTTAAACATTCCTCAACCTTGCTACATGAGTAATTCATCTGAGAG 48466
QY 966 AAACCATTCATTTTCACTGTGTGTGTAAGGGGTTAATTCGACAAACACTTGAAAAAG 1025
DB 48465 AAACCTACCAATGTGAAGAAATGTGTAAAGCCTTTAAACAGTCTCAACCTTACTAGA 48406
QY 1026 CATGAATATCCCATACAAAGTCAATTTAAATGTACATTTGAAAAATTTGTCAAGACATTT 1085
DB 48405 CATAGAGATTCATATCTGAGAGAAACCTACCAATGTGAAAAATGTGCAAAAGCTTTT 48346
QY 1086 TATTAACATCAATCTTAA---GACATCATATATTTATCTGTATGAAAAAACATTAACG 1142
DB 48345 AACCATGCTCAAACTTACTGACATTAAGAAATTTATCTGTGTGAAGAACTCTACAA 48286
QY 1143 TGTAAACATGTATTAATTAAGTTTTCATCTGACCTTCANAAATTTAGCAACATTAATTA 1202
DB 48285 CTTAAAGATGTAAACATGTATTTGAAAAACATTCANAAAGTTTCTAAACATTAAGAAAT 48226
QY 1203 CATCATGTGTGAGTCTCTGCTTATCAATGTGATCATCTGCTTTTAAAAATTTCCAA 1262
DB 48225 TATGCTGTGTGAAGATCTTGAAGATGTGAAGAAATGTGAAGAAACCTTTAAAGTTGTAC 48166
QY 1263 ACTGTGTCAGTATTAATTCATTAATTAACATGATGATCCAAACT 1309
DB 48165 ACTGTATTTGATTAAGTATTAATTCATTAATTAAGTATTAAGT 48119

RESULT 14
AC138469/c 169500 bp DNA linear HTG 08-JAN-2003
LOCUS Homo sapiens chromosome 19 clone RP11-189C24, *** SEQUENCING IN
AC138469
AC138469.1 GI:27544967
AC138469
HTG: HTGS_PHASE2; HTGS_ACTIVIFIER.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 169500)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
Unpublished
2 (bases 1 to 169500)
DOE Joint Genome Institute and Stanford Human Genome Center.
Submitted (08-JAN-2003) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
* NOTE: This is a "working draft" sequence. It currently
* consists of 5 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 75796: contig of 75796 bp in length
* 75797 75896: gap of unknown length

```

```

* 75897 138071: contig of 62175 bp in length
* 138072 138171: gap of unknown length
* 138172 157899: contig of 19728 bp in length
* 157900 157999: gap of unknown length
* 158000 166058: contig of 8059 bp in length
* 166059 166158: gap of unknown length
* 166159 169500: contig of 3342 bp in length.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web Site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: RP11-189C24
Center clone name: RP11-189C24
-----
Summary Statistics
Consensus quality: 16586 bases at least Q40
Consensus quality: 16581 bases at least Q30
Consensus quality: 166156 bases at least Q20
Estimated insert size: 169095000; agarose-fp estimation
Estimated insert size: 166687; sum-of-contigs

estimation
Quality coverage: 0 in Q20 bases; agarose-fp
estimation
Quality coverage: 1 in Q20 bases; sum-of-contigs
location/Qualifiers
1.169500
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="19"
/clone="RP11-189C24"

ORIGIN
Query Match 4.8%; Score 99.4; DB 2; Length 169500;
Best local similarity 50.6%; Pred. No. 1.1e-06;
Matches 297; Conservative 0; Mismatches 281; Indels 9; Gaps 2;

QY 726 GAAAGTACGAAACCAATGATATCATCTTTAATATCTTCTTCATCAGCTCC 785
DB 49667 GAATGTGGCAAAAGCTTTTAAACGATCCTCAATCTTATCAACATTAAGAAATTCATCT 49608
QY 786 AAAAATTAATTTTGCACATATGAAGGGTGTGATAAGCCCTATATGACCATCATTTA 845
DB 49607 GAGAGAAATCCTACAAATGTGAAGATGTGGCAAGCTTCTCATGATCTCTCAAACTT 49548
QY 846 GAGCAACATTTTAAAGAACCAAGTAAATGATGACCGTATTAATGTACAGTGAAGATGT 905
DB 49547 ACTGAAATTAAGAAATTCATATCTGAGAGAAACCTACACATGT-----GAAAGATGT 49494
QY 906 GATTAAGCATTTTTCAGAAAAATGACATTTTGAAGAACATATTTGATACATTCGCAAAA 965
DB 49493 GGCAGAGCCCTTAAACATTCCTCAACCTTGCTACATTAAGTATTAATTCATCTGAGAG 49434
QY 966 AAACCATTCATTTTCACTGTGTGTGTAAGGGGTTAATTCGACAAACACTTGAAAAAG 1025
DB 49433 AAACCTACCAATGTGAAGAAATGTGTAAAGCCTTTAAACAGTCTCAACCTTACTAGA 49374
QY 1026 CATGAATATCCCATACAAAGTCAATTTAAATGTACATTTGAAAAATTTGTCAAGACATTT 1085
DB 49373 CATTAAGAAATTCATCTGAGAGAAACCTACCAATGTGAAAAATGTGCAAAAGCTTTT 49314
QY 1086 TATTAACATCAATCTTAA---GACATCATATATTTATCTGTATGAAAAAACATTAACG 1142
DB 49313 AACCATGCTCAAACTTACTGACATTAAGAAATTCATCTGAGAGAAATCTACAA 49254
QY 1143 TGTAAACATGTATTAATTAAGTTTTCATCTGACCTTCANAAATTTAGCAACATTAATTA 1202
DB 49253 CTTAAAGATGTAAACAGATTTTGAAGAAACATTCANAAAGTTTCTAAACATTAAGAAAT 49194
QY 1203 CATCATGTGTGAGTCTCTGCTTATCAATGTGATCATCTGCTTTTAAAAATTTCCAA 1262

```

```

Db      49193  TATCTGCTGAGTAATCTTGAAGATGGAAGATGACAAACCTTTAAAGTTGTAC 49134
QY      1263  ACTGGCAGTATTCATTTCAATTAACACTGCATCCAAACT 1309
Db      49193  ACTGATGTGCAATAGATTAATCATCTAAAAAACCCTATAGT 49087

RESULT 15
AKI22869
LOCUS   AKI22869
DEFINITION Homo sapiens cDNA FLJ16502 fis, clone FEBRA2006664, moderately similar to Zinc finger protein 43.
ACCESSION AKI22869
VERSION   AKI22869.1 GI:34528067
KEYWORDS  oligo capping; fis (full insert sequence).
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
REFERENCE 1
AUTHORS   Tanigami,A., Fujiwara,T., Shibahara,T., Goto,Y., Hirao,M., Shimizu,F., Wakebe,H., Ono,T., Hishigaki,H., Watanabe,T., Ozaki,K., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saico,K., Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Nagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahara,K., Masuho,Y., Nagai,K. and Isogai,T.
TITLE     NEDO human cDNA sequencing project
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 2237)
AUTHORS   Isogai,T. and Yamamoto,J.
TITLE     Direct Submission
JOURNAL   Submitted (15-JUL-2003) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.
FEATURES
Source    1..2237
           /organism="Homo sapiens"
           /mol_type="mRNA"
           /db_xref="taxon:9606"
           /clone="FEBRA2006664"
           /tissue_type="brain"
           /clone_id="FEBRA2"
           /dev_stage="fetal"
           /note="cloning vector: pME18SFL3"

ORIGIN
Query Match      4.8%; Score 98.2; DB 9; Length 2237;
Best Local Similarity 52.0%; Pred.No.3.6e-06;
Matches 273; Conservative 0; Mismatches 243; Indels 9; Gaps 2;

QY      788  AAAGTAAATTGACATATGAGAGGTGTGATAAAGCTATATATGACCATATTATTAGA 847
Db      1652  AGAGAAATCTTACAAATGTGAAGATGTGGCAAGCTTTCTATGATCTCAAAAATTAC 1711
QY      848  GCAACATTTAAGAACCCAGATATGATCGACCGTATPAATGTACAGTGACGATTGTGA 907
Db      1712  TGAACATAAGAAATTTATCTGAGAGAAACCCTTACACATGT-----GAAAGATGTGG 1765
QY      908  TAAAGCAATTTTCAAGAAATCACTTTGGAACACATATTTGATCACTTCGAAAAAAA 967
Db      1766  CAAAGCCTTAACCATCTCACACCTTGTCTACACATAGGTAAATTCACTCTGAGAGAA 1825

```

```

QY      968  ACCATTCATTTGTCAGTGTGTGTAAGGGGTTAATTCGACACACTTGAAGACA 1027
Db      1826  ACCCTACATGTGAAGAAATGTGTAAAGCCTTTAACAGTCTCTACACCTTCTACACA 1885
QY      1028  TGAATCACCCTATACAAAGTCAATTAATGTACATTTGAAAATTTGCAAGACATTTTA 1087
Db      1886  TAAAGAAATTCATATCTGAGAGAAACCCCTACCATGTGAAAANTGTGGCAAAGCTTTAA 1945
QY      1088  TAAACATCAATCTTTAA---GACATATATATATCTGTTCATGAAAAAAACATTAAGTG 1144
Db      1946  CAGTCTCCAAACCTTACATGACATTAAGAAAATTCATAGTGAGAAAACCTTACAAACC 2005
QY      1145  TAAACATGTATTAAGTTTTCATCTGACCTTAATAAATTAAGCAACATTAATTAACA 1204
Db      2006  TAAAGATGTAAAGATTTTGAACACCTTCAAAAGTTTCTTAAACATTAAGAAATTA 2065
QY      1205  TCATGTGATCTCTGCTTATCAATGTATCATCTCTGTTGTTTAAAAATTTCCAAAC 1264
Db      2066  TGCTGTGAGAAATCTTAGAAATGTGAAGAAATGTAACAAAACCTTTAAAGTTGTACAC 2125
QY      1265  TTGTCAGATTTTCAATTTTCATATTAACCACTGCATCCAAACT 1309
Db      2126  TTGATTTGCATTAAGATTAATTCATCTAAAAAACCCTATAGT 2170

```

Search completed: July 28, 2004, 17:19:28
Job time : 8124 secs

OM nucleic - nucleic search, using sw model

(without alignments)

11368.710 Million cell updates/sec

Title: US-09-831-804-1
Perfect score: 2060
Seminar: 1-2-2001

Scoring table: IDENTITY_NUC

searched: 27513289 seqs, 14931090276 residues
Total number of hits satisfying chosen parameters: 55026578

| | |
|------------------------|-------------|
| Maximum DB seq length: | 20000000000 |
|------------------------|-------------|

```

post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

```

1 em_estpa:*
2 em_estlun:*
3 em_estln:*
4 em_estlun:*
5 em_estlcv:*
6 em_estpl:*
7 em_estro:*
8 em_hlc:*
9 gb_estl:*
10 gb_est2:*
11 gb_hlc:*
12 gb_est3:*
13 gb_est4:*
14 gb_est5:*
15 em_estfun:*
16 em_estom:*
17 em_gss_jum:*
18 em_gss_inv:*
19 em_gss_pln:*
20 em_gss_vrt:*
21 em_gss_fun:*
22 em_gss_mam:*
23 em_gss_mus:*
24 em_gss_pro:*
25 em_gss_rod:*
26 em_gss_phg:*
27 em_gss_vrt1:*
28 gd_gss61:*
29 gb_gss62:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result | Score | Query | Length | DB | ID | Description |
|--------|-------|-------|--------|----|----------|--------------------|
| No. | | | | | | |
| c 1 | 182.2 | 8.8 | 628 | 28 | B2298000 | B2298000 C3957.F |
| c 2 | 167.6 | 8.1 | 959 | 29 | AK0060MP | AL405671 T7 end of |
| 3 | 98.2 | 4.8 | 1609 | 11 | AK032220 | AK032220 Mms musci |
| 4 | 97.8 | 4.7 | 2647 | 11 | BC047646 | BC047646 Homo sapi |

| | | | | | | |
|---|----|------|-----|------|----|--------------|
| C | 5 | 97.6 | 4.7 | 745 | 9 | AU123448 |
| C | 6 | 92.4 | 4.5 | 1101 | 29 | CNS003498 |
| C | 7 | 92.4 | 4.5 | 2672 | 11 | BC036394 |
| C | 8 | 92 | 4.5 | 591 | 28 | AQ005136 |
| C | 9 | 91 | 4.4 | 593 | 28 | B99387 |
| C | 10 | 91 | 4.4 | 1163 | 13 | BX414627 |
| C | 11 | 90.4 | 4.4 | 1253 | 13 | BC028252 |
| C | 12 | 90.4 | 4.4 | 2694 | 11 | BC025227 |
| C | 13 | 90.4 | 4.4 | 2698 | 11 | BC037782 |
| C | 14 | 89.2 | 4.3 | 604 | 10 | BE161630 |
| C | 15 | 89.2 | 4.3 | 3609 | 11 | BC032590 |
| C | 16 | 88.6 | 4.3 | 535 | 9 | AL705393 |
| C | 17 | 88.6 | 4.3 | 2174 | 11 | AK033001 |
| C | 18 | 88.6 | 4.3 | 2330 | 11 | BC037426 |
| C | 19 | 88.6 | 4.3 | 2622 | 11 | AK033958 |
| C | 20 | 88.4 | 4.3 | 1084 | 13 | BX456765 |
| C | 21 | 88 | 4.3 | 895 | 13 | BUL195416 |
| C | 22 | 87.6 | 4.3 | 832 | 13 | BX437291 |
| C | 23 | 87.6 | 4.3 | 1201 | 13 | BX355654 |
| C | 24 | 87.4 | 4.2 | 1200 | 13 | BX437758 |
| C | 25 | 87.4 | 4.2 | 1392 | 29 | CG377503 |
| C | 26 | 87.2 | 4.2 | 579 | 28 | AQ347265 |
| C | 27 | 87.2 | 4.2 | 687 | 28 | AQ369266 |
| C | 28 | 87.2 | 4.2 | 907 | 13 | BQ433752 |
| C | 29 | 87 | 4.2 | 660 | 14 | CD770140 |
| C | 30 | 87 | 4.2 | 842 | 13 | BUL507408 |
| C | 31 | 86.8 | 4.2 | 521 | 28 | AQ455626 |
| C | 32 | 86.8 | 4.2 | 705 | 28 | AQ194282 |
| C | 33 | 86.8 | 4.2 | 860 | 28 | AQ749175 |
| C | 34 | 86.8 | 4.2 | 897 | 13 | BQ431141 |
| C | 35 | 86.8 | 4.2 | 1906 | 11 | AK013043 |
| C | 36 | 86.6 | 4.2 | 694 | 12 | BG656505 |
| C | 37 | 86.6 | 4.2 | 1098 | 13 | BX377526 |
| C | 38 | 86.4 | 4.2 | 874 | 13 | BUL171453 |
| C | 39 | 86.4 | 4.2 | 2791 | 11 | BC020045 |
| C | 40 | 86.2 | 4.2 | 523 | 14 | CA874049 |
| C | 41 | 86 | 4.2 | 557 | 10 | BP817669 |
| C | 42 | 86 | 4.2 | 657 | 14 | CP135920 |
| C | 43 | 86 | 4.2 | 685 | 29 | AG149704 |
| C | 44 | 86 | 4.2 | 801 | 14 | CD657347 |
| C | 45 | 85.8 | 4.2 | 1201 | 13 | BX458623 |
| | | | | | | AU123448 |
| | | | | | | AU123448 |
| | | | | | | AL063921 |
| | | | | | | Drosophila |
| | | | | | | Homosapiens |
| | | | | | | Homosapiens |
| | | | | | | CITR-HSP- |
| | | | | | | AG005136 |
| | | | | | | B99387 |
| | | | | | | CITR-HSP-222 |
| | | | | | | BX414627 |
| | | | | | | BX414627 |
| | | | | | | BC028252 |
| | | | | | | Mus musculus |
| | | | | | | BC025227 |
| | | | | | | Homosapiens |
| | | | | | | BC037782 |
| | | | | | | Homosapiens |
| | | | | | | BE161630 |
| | | | | | | MR3-HT04 |
| | | | | | | BC032590 |
| | | | | | | Homosapiens |
| | | | | | | AL705393 |
| | | | | | | DKFP2686M |
| | | | | | | AK033001 |
| | | | | | | Mus musculus |
| | | | | | | BC037426 |
| | | | | | | Homosapiens |
| | | | | | | AK033958 |
| | | | | | | Mus musculus |
| | | | | | | BX456765 |
| | | | | | | BX456765 |
| | | | | | | BUL195416 |
| | | | | | | AGENCOURT |
| | | | | | | BX437291 |
| | | | | | | BX437291 |
| | | | | | | BX355654 |
| | | | | | | BX355654 |
| | | | | | | BX437758 |
| | | | | | | BX437758 |
| | | | | | | CG57503 |
| | | | | | | P052-4-CO |
| | | | | | | |

ALIGNMENTS

| | |
|------------|---|
| RESULT 1 | BZ29800/c |
| LOCUS | BZ298000 628 bp DNA linear GSS 31-OCT-2002 |
| DEFINITION | Cg33957.f1 Candida glabrata Random Genomic Library Candida glabrata genomic clone Cg33957, genomic survey sequence. |
| ACCESSION | BZ298000 |
| VERSION | BZ298000 |
| KEYWORDS | BZ298000.1 GI:24440936 |
| SOURCE | GSS. |
| ORGANISM | Candida glabrata Candida glabrata Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; mitosporic Saccharomycetales; Candida. 1 (bases 1 to 628) |
| REFERENCE | Wong, S., Fates, M.A., Zimmermann, W., Butler, G. and Wolfe, K.H. Evidence from comparative genomics for a complete sexual cycle in the 'asexual' pathogenic yeast Candida glabrata |
| AUTHORS | JOURNAL |
| TITLE | MEDLINE |
| JOURNAL | PUBMED |
| MEDLINE | 12620120 |
| PUBMED | Contact: Wong S Department of Genetics, Smurfit Institute Trinity College Dublin Dublin 2, Ireland Tel: 353 1 6082319 Fax: 353 1 6798558 Email: swong@tcd.ie |
| COMMENT | Class: plasmid ends. |

FEATURES

Location/Qualifiers
1. 628
/organism="Candida glabrata"
/mol_type="genomic DNA"
/strain="CBS 138"
/db_xref="taxon:5478"
/clone="CG3957"
/clone_1ib="Candida glabrata Random Genomic Library"

ORIGIN

Query Match 8.8%; Score 182.2; DB 28; Length 628;
Best Local Similarity 59.5%; Pred. No. 4.7e-18;
Matches 348; Conservative 0; Mismatches 228; Indels 9; Gaps 2;

QY ATATGAGGGGTGTGATTAAGCCCTTAATGACACCATTAATTAGACAACTTTAAGAAC 862
DB ATACGATTAATTTGACAGAGCCCTTTACAGAGCCGCTCTCTACAGAAACACAGAAATAC 566
QY 625 ATACGATTAATTTGACAGAGCCCTTTACAGAGCCGCTCTCTACAGAAACACAGAAATAC 566
QY 863 CCAAGATGATGACGACCGTAAATGTAAGTGAAGATTTGATTAAGCAATTTTTCAG 922
DB 565 ---AGTGCATCTTGGGAGAAACCTTGAATGTAAATGTAAGTGAAGTTCATTTACTAA 509
QY 923 AATATCATTGTTGGAACACATATTTGATGACATTCGGAACAAAAACCATTCATTGTTTC 982
DB 508 AAGGATCCATTGAGAGACACTTATACACACAGACAGATGAAGACCGTTTATTTGTTTC 449
QY 983 AGTGTGTGTAAGGGGTAAATTTCTGACAAACCTTGAAGAACATGAATACCCCATAC 1042
DB 448 ATTTTGGGAGAGGGGCTAAATTAATCTAGCAACAACTGAAGACATGAGTTAACTCACAC 389
QY 1043 AAGATCATTAATATGTCATTTGAAATTTGTCAGAAAGCATTTTATTAACATCAATCTTT 1102
DB 388 CAATTCCTTCAATGTGAATATGAAGGTGTGAATGAAGTTTCTAAGACCCCTCATAT 329
QY 1103 AAGACATCATATATTTATCTGTTTCATGAAGAAACATTAACGTAAACATGTATTAAGT 1162
DB 328 AAGGGCTCAATTTTACGAGTTCTTTACAAAGCTTAATATGTACAGATGAACAAAG 269
QY 1163 TTTTCATCGACCTTCANAAATTTAGCAACATTAATTAACATGATGGTGCATCTCC--- 1219
DB 268 TTTTCAAGACCTTACAGACTCAAAATTCACATGACCAACACATATCCAGATGTTGT 209
QY 1220 ---TGTATTAATGATGATCATCTGCTGTTGTTTAAATTTCCAACTGGTCAGATTT 1276
DB 208 CAATGCTTATCAATGTACCTTCAAGTGTGTTCTTAAGGTTTCAAAAGATGTCGCTT 149
QY 1277 ACAATTCATATTAACAACTGACATCCAAACTTAATATGTCCTTAATGTGTTAAAGTTG 1336
DB 148 AAGATTACATGTTAAATATGATCATCCGAAGTTAAATATGCGCTATATGCAATTAACCTG 89
QY 1337 TGTGGGAAAAAAGTTTATCTTCAATATGTTAAAGTCATGATGA 1381
DB 88 TGTGGGGAAGCGGTTTAAATATGACATGAATAATCCAGATGA 44

RESULT 2
CNS06MMP/c 959 bp DNA linear GSS 17-JUN-2001
LOCUS T7 end of clone AU0AA005F10 of library AU0AA from strain CBS 3082
DEFINITION of Saccharomyces kluyveri, genomic survey sequence.
ACCESSION ALA05671.1 GI:12168715
VERSION ALA05671.1 GI:12168715
KEYWORDS GSS.
SOURCE Saccharomyces kluyveri
ORGANISM Saccharomyces kluyveri
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes;
1 (bases 1 to 959)
Souchier,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
deMontigny,U., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
Malpertuy,A., Neuvéglise,C., Olier-Kalogeropoulos,O., Potier,S.,
Sautin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,

TITLE

Wincker,P. and Weissenbach,J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
FEBS Lett. 487 (1), 3-12 (2000)

MEDLINE

20584711
1152876

PUBMED

2 (bases 1 to 959)
Neuvéglise,C., Bon,E., Lepingle,A., Wincker,P., Artiguenave,F.,
Gallardin,C. and Casaregola,S.
Genomic exploration of the hemiascomycetous yeasts: 9.
Saccharomyces kluyveri
FEBS Lett. 487 (1), 56-60 (2000)

REFERENCE

AUTHORS

1152884

JOURNAL

20584719
1152884

MEDLINE

3 (bases 1 to 959)
Genoscope.
Direct Submission
Submitted (07-SEP-2000) Genoscope - Centre National de Séquençage,
2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
secrefgenoscope.cns.fr - Web : www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazii, Zygosaccharomyces rouxi,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.

COMMENT

AUTHORS

1152884

JOURNAL

20584719
1152884

MEDLINE

3 (bases 1 to 959)
Genoscope.
Direct Submission
Submitted (07-SEP-2000) Genoscope - Centre National de Séquençage,
2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
secrefgenoscope.cns.fr - Web : www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazii, Zygosaccharomyces rouxi,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.

FEATURES

source

1. 959
/organism="Saccharomyces kluyveri"
/mol_type="genomic DNA"
/strain="CBS 3082"
/db_xref="taxon:4934"
/clone="AU0AA005F10"
/clone_1ib="AU0AA"
/note="end : T7"
complement (<71..>958)
/note="similar to Saccharomyces cerevisiae ORF YPR186c [PZ1 ; TFI1A (transcription initiation factor)]"
/evidence=not_experimental

ORIGIN

Query Match 8.1%; Score 167.6; DB 29; Length 959;
Best Local Similarity 57.2%; Pred. No. 5.7e-16;
Matches 324; Conservative 1; Mismatches 235; Indels 6; Gaps 1;

QY 930 CATTGGAACACATATTTGATATCAATTCGAAAAAAACCATTCATTGTTCAAGTGT 989
DB 958 CACTTGGAAGACATATTTGATATCAATTCGAAAAAAACCATTCATTGTTCAAGTGT 899
QY 990 GGTAAAGGGGTTAATTTCTGCAACACCTTGAAGAAAGCATGAATACCATTAACAAGTCA 1049
DB 898 GGTAAAGGGGTTAATTTCTGCAACACCTTGAAGAAAGCATGAATACCATTAACAAGTCA 839
QY 1050 TTTAATGTACATTTGGAATATTTGTCAGACACATTTTAAACATCAATCTTTAAGACAT 1109
DB 838 TTCAATGTGAATATGAGGGCTGCAAGCAATTTTTCAGAGATTCGCACTAAGTCA 779
QY 1110 CATATATTATCTGTTTCATGAAAAAAACATTAAGTGAACCAATGATTAAGTCTTCACT 1169
DB 778 CATACACTTGGTATGACCTTGAGAAATTTGACCTGTGAGCATTTGGGTTAAATGTTCCAA 719
QY 1170 GCAGCTTAAATTTAGCAACACATTAATTAAGCAATCTGATGATCTCCCTGCT----- 1223
DB 718 AGGCATATATCTGTCAGAAACCATCTTCAAGAGATTCGCAATGATGATTTAAACGAAA 659
QY 1224 TATCAATGTGATCATCCCGTGTGTTTAAATTTCCAACTTGGTCAGTATTAACATTT 1283
DB 658 TATCAATGTGATCATCCCGTGTGTTTAAATTTCCAACTTGGTCAGTATTAACATTT 599

QY 1284 CATATAAACAAGTCCATCCAAACTTAATGCTTAATGCTTAAGTGTGTTGGG 1343
 Db 598 CATTCAGGCGGACCATCCCAAGTTGACGTGACGTATGTGGCAAGCTTGTTGGT 539
 QY 1344 AAAAAAGTTTATTCATCATATGTTAAGTCATGATGTTTACCATGATCAAAATANG 1403
 Db 538 GAGCTGTGTCTACGACATGACATCGAGTCCATATCAAGGCTTTAGTAATTAATAATTGG 479
 QY 1404 ACTTGTCATTTATTTGTGATGTGGGAAATTTGCAAGAAAATGAATTAAGTGAATTTAT 1463
 Db 478 AAATGTACCAATTTGTCTAGCAAGTCTCTTTGCTTAAGAAAGGTGATCTTTATCTCATAC 419
 QY 1464 AATATCTTCATGATGTTAATATCC 1489
 Db 418 ATGGAACATCATTAAGATGATATACC 393

RESULT 3
 AK032220
 LOCUS
 DEFINITION Mus musculus adult male olfactory brain cDNA, RIKEN full-length
 accession AK032220.1 GI:26328050
 VERSION AK032220
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 AUTHORS Carninci, P. and Hayashizaki, Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636

REFERENCE
 AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Komno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20499374
 PUBMED 11042159

REFERENCE
 AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Komno, H., Akiyama, J., Nishi, K., Katsunai, T., Taashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujiwake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
 TITLE RIKEN integrated sequence analysis (RISA) system-384-format
 sequencing pipeline with 384 multicapillary sequencer
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
 MEDLINE 20530913
 PUBMED 11076861

REFERENCE
 AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
 PANTOM Consortium.
 TITLE Functional annotation of a full-length mouse cDNA collection
 JOURNAL Nature 409, 685-690 (2001)
 PUBMED 11076861

REFERENCE
 AUTHORS The PANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
 TITLE Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 JOURNAL Nature 420, 563-573 (2002)
 PUBMED 11076861

REFERENCE
 AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hirao, K., Hirose, T.,
 Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,

COMMENT

Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Komno, H., Kouda, M.,
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
 Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
 Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
 Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
 Muramatsu, M. and Hayashizaki, Y.
 TITLE Direct Submission
 JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)
 CDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Please visit our web site for further details.
 URL: http://genome.gsc.riken.go.jp/
 URL: http://fantom.gsc.riken.go.jp/
 Location/Qualifiers

FEATURES

source

1. 1609
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM_DB:6430503015"
 /db_xref="WGI:2395676"
 /db_xref="taxon:10090"
 /clone="6430503015"
 /sex="male"
 /tissue_type="olfactory brain"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="adult"
 330..1223
 /note="unnamed protein product; MSZF33 (FRAGMENT) homolog
 [Mus musculus] (SPRR|088251, evidence: PASTY, 87.1%ID,
 100%length, match=255)
 putative"

CDS

/codon_start=1
 /protein_id="BAC27766.1"
 /db_xref="GI:26328051"
 /translation="MLETFRYNNANVGMEDSNIEERCESSRRHGRHRTGKPYE
 GTOGGAEPYHSHSLQMRKLIHTGKRYKNCQDKAYSRHSIIQIHRHTSGEPYCN
 QCGKAPVYSHLKHVHTGTGKPYKNCQGRKFAFACNKLQKERTHTGKPYKNCQ
 NAFVYESTLQVHKHTGKPYKNCQGRKFAFARHSHLKVHKIHTGKPYKNCQCGK
 ALAYSTGLQVHRTGKPYKNCQGRKFAFARHSHLKVHKIHTGKPYKNCQCGKAF
 VSSDLKRHRVHTGKPYKNCQGRKFAFARHSHLKVHKIHTGKPYKNCQCGKAF
 1594..1599
 /note="putative"
 1609
 /note="putative"
 /note="putative"

ORIGIN

Query Match 4.8%; Score 98.2; DB 11; Length 1609;
 Best Local Similarity 51.2%; Pred. No. 1.1e-05;
 Matches 310; Conservative 0; Mismatches 283; Indels 12; Gaps 3;

QY 779 AGCTCCCAAAAGTATATTGGACATATAGAGGTTGTAAGCTTATATGACCATC 838
 Db 521 AACTACTGAGAGAAAGCGTACAAATGCAATCATGTGTAAGCTTATACGACACAG 580
 QY 839 ATTATTAGGCAACATTTAAGAACCCACGATATGATGACCGTATTAATGTCAGTGA 898
 Db 581 CATTTACAAATACATTAAGAAACACATGTGGAGGAACCTATGAATGTA-----A 634
 QY 899 CGATTGTGTAAGACATTTTTCAGAAATCACATTTGGAAGACATATGTAACATTC 958
 Db 635 TCATATGTGTAAGCGCTTTACACACACGATCTTCAAAATACATATGTTACACATAC 694

QY 959 CGAAAAAACCATTCATCTTCTGAGTGTGTGTAAGGGGTAAATCTCTGACAACTT 1018
 DB 695 TGGAGAGAACCTTACAAATGTAATCAATGTGTAAAGCCTTTCAGTCTAATAACT 754
 QY 1019 GAAAGACATGAATCCACCAATCAAAAGTCATTTAAATGATCAATTTGATTCAGA 1078
 DB 755 TCAAAAACATGAAGAACACATCTGAGAGAAACCTTAGAAGTGTATCAATTAATA 814
 QY 1079 AGCATTT--TATAACATCAATCTTTAAGATCATATATTTATCTGTCTATGAAAAAC 1135
 DB 815 AGCCTTGTATATGAAAGTATTTATTAACAAGTTCATPAAAAAACACATCTGAGAGAAACC 874
 QY 1136 ATTAACGTTAAACAAAGTAAATTAAGTTTTCATCTGACCTTCAAAATAGACAACATA 1195
 DB 875 TTACAAATGTATATGATGTGTAAAGCCTTTCACACACAGTCAATCTCAAAAGTCATTA 934
 QY 1196 ATTAACATCATGATGATCTCTGCTTATCAATGTATCATCTGTTGTTTAAAAA 1255
 DB 935 AATAACATCATGAGAGAAAC--CTTACAAATGTATCAATGTGTAAAGCCTTGC 991
 QY 1256 TTTCCAACTGTGTGATTTCAATTTCAATTAATAAACTGATCCAAAATTAATG 1315
 DB 992 ATATCATGTACTCTCCAAAGTACATCAAAAGACATCTGAGAGAAAGCCCTATGATG 1051
 QY 1316 TCTTAATGTGTAAAGTGTGTGTTGGAAAAAGTTATCTTCAATATGTTAAGTCA 1375
 DB 1052 TGAGCAATGTGGCAAGCCTTTCGAAATCAAAATTTATTTCCAAAGTACATAAAGATACA 1111
 QY 1376 TGATG 1380
 DB 1112 TACTG 1116

RESULT 4
 LOCUS BC047646 2647 bp mRNA linear HTC 19-NOV-2003
 DEFINITION Homo sapiens zinc finger protein 85 (HPF4, HTF1), mRNA (cDNA clone IMAGE:4828290), with apparent retained intron.
 ACCESSION BC047646
 VERSION BC047646.1 GI:29126804
 KEYWORDS HTC.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 2647)
 AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Ditchkenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, J.L., Scheet, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S., Lochuano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hult, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butcherfield, Y.S., Krzyzanski, M.I., Skalek, U., Smolins, D.E., Schermer, A., Schein, J.E., Jones, S.J. and Marra, M.A.
 TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 MEDLINE 22388257
 PUBMED 12477932
 REFERENCES 2 (bases 1 to 2647)
 AUTHORS Strausberg, R.
 TITLE Direct Submission
 JOURNAL Submitted (03-MAR-2003) National Institutes of Health, Mammalian

REMARK

Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Miklos Palokovics, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shitaki Toshiyuki and Piero Carninci (RIKEN)
 DNA Sequencing By: The I.M.A.G.E. Consortium (LBNL)
 Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www.shgc.stanford.edu>
 Contact: (Dickson, Mark) mdcpaxil.stanford.edu
 R. M.
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,

FEATURES

source

1. 2647
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4828290"
 /tissue_type="Testis"
 /clone_id="NIH MGC_97"
 /lab_host="DH10B"
 /note="Vector: pBluescript"

ORIGIN

Query Match 4.7%; Score 97.8; DB 11; Length 2647;
 Best Local Similarity 50.1%; Pred. No. 9.4e-06;
 Matches 330; Conservative 0; Mismatches 317; Indels 12; Gaps 3;
 QY 726 GAAAGTACGAAACCAATCATATCTTAATCTCTCTTCAATCAAGTCC 785
 DB 1235 GAATGTGGCAAAACCTTTAAACGATCTGCAACTCTTACCCATAGATATTCATCT 1294
 QY 786 AAAAGTATATTGGACATATGAAGGGTGTATGAAGGCTTAATGACCATATTATTA 845
 DB 1295 GGAGGAAACCTTCAATGTAAAGAAATGTGTAAAGCTTTAAACCGATCTTCAACCTT 1354
 QY 846 GAGCAACATTTAAGAACCCACATAGATATGACCGTATATGACGTGAGAGATGT 905
 DB 1355 ACTACCCATTAAGAAATTCATATCTGAGAGAAACCTTACAAATGT-----GAAGATGT 1408
 QY 906 GATAAGCATTTTTCAGAAATCATTGGAAACACATATGTATCAATTCGAAAAA 965
 DB 1409 GGCAAGCCTTTAAGACATCTCTCAAACTTACTACAGATAGATATCATATCTGAGAG 1468
 QY 966 AAACCATTCATGTGTCAAGTGTGTGAAGGGGTTAATTTCCGAAACATCTTGAAGA 1025
 DB 1469 AAACCTTCAATATGTAAGAAATGTGAAAGCCTTTAAACAGTCTGACACCTTACCA 1528
 QY 1026 CATGAATCACCCATACAAAGTCAATTAATGTATCAATTTGAATTTGCAAGAACATTT 1085
 DB 1529 CATGAGTAATTCATACGAGAGAAACCTTACAAATGTGAATTTGGAAGAACCTTT 1588
 QY 1086 TATAA--ACATCATCTTTTAAGACATATATATCTGTTCAATGAAAAACATTAACG 1142
 DB 1589 AATCATTTCTGACACCTTACATACATATTAATTAATTAATTAATTAATTAATTA 1648
 QY 1143 TGTAAACATGTATTAAGTTTCACTGACCTTAATTAATTAATTAATTAATTAATTA 1202
 DB 1649 TGTAAAGATGTGTGAAGCTTTTAACATCTTCAACCTTCAATTAATTAATTAAT 1708
 QY 1203 CATCATGTGTGATCTCCGTCTTATCAATGTATCATCTCGTGTGTTTAAATTTCCAA 1262
 DB 1709 CATCTGAGAGAGAGCCTTACAAATGTAAAGATGTGAAGAGCTTTTAACCAATCTCT 1768

```

QY      1263  ACTTGTCAGTATTACATTTATATATAAACACCTGCATCCAAAACCTTAATGTCCTTAA 1322
Db      1769  AAACCTTACTGAACATTAAGAAAATTCATCTGAGAGAAACCTTA---TGAATGTGAAAAA 1825
QY      1323  TGTGGTAAAGGTGTGTGGGAAAAAAGGTTTATCTTCACATATGTTAAGCAGATGA 1381
Db      1826  TGTGGCAAAAGCTTTTAACCAAGTCTCTCAATCTTACTAGACATTAAGAAAAGTCATACAGA 1884

RESULT 5
AUI23448      745 bp  mRNA  linear  EST 01-AUG-2002
LOCUS      AUI23448
DEFINITION  AUI23448 NT2RM2 Homo sapiens cDNA clone NT2RM2000319 5', mRNA
ACCESSION  AUI23448
VERSION    AUI23448.1
KEYWORDS   GI:10948164
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 745)
AUTHORS   Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,
Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and
Iisogai,T.
TITLE      HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,
Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T.,
Suzuki,Y., Sugano,S., Iisogai,T.)
JOURNAL    Unpublished (2000)
COMMENT    Contact: Takao Iisogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project: 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
    source
        1..745
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="NT2RM2000319"
            /cell_type="leukocarcinoma"
            /cell_line="NT2"
            /clone_lib="NT2RM2"
            /note="Vector: PME18SFL3; mRNA from uninduced NT2 neuronal
            precursor cells"
ORIGIN
Query Match      4.7%; Score 97.6; DB 9; Length 745;
Best Local Similarity 50.2%; Pred. No.2.1e-05;
Matches 328; Conservative 0; Mismatches 314; Indels 12; Gaps 3;

QY      726  GAAAGTACGAAACCAATCGATATCATCTTTAATATCTTCTTCATCACTGCC 785
Db      87  GAATGGGCAAACTTTTAACCGATCTCAACTCTTACACCAATAGATTAATCTACT 146
QY      786  AAAAATATATTTGCACATATGAGGGGTGTGATAAGCCTATATATGACCATATTATTA 845
Db      147  GAGAGAAAACCTTACCAATGTAAAGATGTGTAAGCTTTAAACGTTCTTCAACCTT 206
QY      846  GAGCAACATTTAAGAACCCACAGTATATGATCGACCGTATTAATGTATGACATGCACTTGT 905
Db      207  ACTACCCATAGAAAATTTATCTACTGAGAGAAACCTTACAAATGT-----GAGGATGT 260
QY      906  GATTAAGCATTTTTCAGAAAATCAATTTGAAACACATATTTGTATCATCTCCGAAAA 965
Db      261  GGCAAGGCTTTAAGCAGTCTCAACCTTACTACACATATAGATTAATCATCTGGAGG 320

```

```

QY      966  AAACATTCATTTGTCAGTGTGTGTAAGGGGTATATTCGACAAACCTTGAAGA 1025
Db      321  AAACCTTCAAAATGTAAATAATGTGAAAAGCCTTTAACAGTGTGCACACTTACTTCA 380
QY      1026  CATGAATATCACCATACAAAGTCATTTAAATGTACATTTGAAAAATGTCAAGACATT 1085
Db      381  CATAGGTAAATTCATCTACTGAGAGAAACCTTACAAATGTGAAAATGTGAAAAGCCTT 440
QY      1086  TATTA---ACATCAATCTTTAAGCATCATATATATCTGTCTTCAAAAAACATTAAG 1142
Db      441  AATCATTTCTCACACCTTACTACACATGAATATTCATCTGAGAGAAACCTTACAAA 500
QY      1143  TGTAAACATATATTAAGTTTTCATCTGACCTCAAAATATAGACAAACATTAATTA 1202
Db      501  TGTAAAGATGTGTAAAGCTTTTAAACACTCTTCAACCTTACTTAACATTAATTAAT 560
QY      1203  CATCATGTGATCTCTCTCTTCAATGATCATCTGATGCTGTTGTTTAAATTTCCAA 1262
Db      561  CATATCTGAGAGAAAGCCTTACAAATGTAAAGAAATGTGAAAAGCTTTTAAACCAT 620
QY      1263  ACTTGTCAGTATTACATTTCAATATTAACAACACTGCATCAAACTTAATGCTTAAA 1322
Db      621  AAACCTTACTGAACATTAAGAAATTCATCTGAGAGAAACCTTA---TGAATGTGAAAA 677
QY      1323  TGTGGTAAAGGTGTGTGGGAAAAAAGTTTATCTTCACATATGTTAAGTCAT 1376
Db      678  TGTGGCAAAAGCTTTTAACCAAGTCTCTCAATCTTACTAGACATTAAGAAAATGCTAT 731

RESULT 6
CNS0039G/c      1101 bp  DNA  linear  GSS 03-JUN-1999
LOCUS      CNS0039G
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION  AL063921
VERSION    AL063921.1
KEYWORDS   GI:4941778
SOURCE     Drosophila melanogaster (fruit fly)
ORGANISM   Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE  1 (bases 1 to 1101)
AUTHORS   Genoscope.
TITLE      Direct Submission
JOURNAL    Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT    Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazuhiro Oosagawa and
Aaron Mammose in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.
FEATURES
    source
        1..1101
            /organism="Drosophila melanogaster"
            /mol_type="genomic DNA"
            /db_xref="taxon:7227"
            /clone="BACR08K10"
            /clone_lib="RPCI-98"
            /note="end : TET3"
ORIGIN

```

| | | | | |
|-----------|------------|---|---|--------|
| | QY | 1369 | TAAATCATGATGATTCTTACCATGATCAAAATATGAGACTTGATTTATTTGGATGTGGCGA | 1428 |
| | Db | 1100 | KASRRGDDTMDRPTRRKDMDWMTKWWTWKKBRADRFRAAGDAPMWDGATWWTA.TWVV | 1041 |
| | QY | 1429 | AATTTCGAAGAAGAAAATGAAATTTAGTTGAACCTTAATAATCTCCAGATGGTATATCC | 1488 |
| | Db | 1040 | WWWMAATWDTWMD.WWWWATPAKTDTATWVMTATWRADVMGRDVGAKGRDDAATTDG | 981 |
| | QY | 1489 | CTGATGATTTTATTAAGAAACCTGAAGTGAAGAAAAATTAGAAGAACCTTATGATCAAGAT | 1548 |
| | Db | 980 | GRRDGGRKDKKDKRKDDGDDKKGKKKKKAATKMWATKMWDMDMDWKXDMKMGDAKOR | 921 |
| | QY | 1549 | CGAATTAATAATTTGCATGAAATTTAGAAACAGAGAAATTAAGGTGAGAGAAATGAG | 1608 |
| | Db | 920 | DDDDGADCKDDGDKGDADDIDGTCKDDDDCKDKMDVKAGTVDATWAAAAADMMWG | 861 |
| | QY | 1609 | AAGATGAAGAAGATAGTCTGATGAAAAAAGAAGTGAATGATGATCAGACTCAATGTCAG | 1668 |
| | Db | 860 | WADDDMTWTPAADADMDWADDWMDAMWMDAMWAGATADRDRMGPRAGRGARRR | 801 |
| | QY | 1669 | CTCAAATCATCAATAAATTCAT--TTACTGCTTCTTTGGAGGTTCAAGAGTGTCTTAA | 1726 |
| | Db | 800 | RKRADDKRDADDBDDDAATTTWTTTTTRDTDMKMKDITWTRWAADBTMRDDDDDR | 741 |
| | QY | 1727 | ACTTATTTCTGAATATGTGGGAAGAAGATCAATTTGCCCTAAGAAATATTTGTATAGAT | 1786 |
| | Db | 740 | AGTARGWRRTFKWRKRWKRRDRMDADADDPARDRRRRGGDAGAGKTKRRRRR | 681 |
| | QY | 1787 | TTCAAGAAATATGATTTAGCTGCACATTTGAATGAGATGATGATTAATTACAAGAT | 1846 |
| | Db | 680 | RATMDRDTAMDADMMWTTTDTDDMDKRRRRKKGARRRRTTBRAAMDWTWKAMWA | 621 |
| | QY | 1847 | TGAGTCAATTTCTTAATATGTATAGAAAAAGAAAGAACTCCAGAAGGTGACATTTGTTAA | 1906 |
| | Db | 620 | KMDKTKTADRDRNAADFTWTDARADRMWAKAPABRARDBARAARDRRWTKGKT | 561 |
| | QY | 1907 | AAAAAGCAGATGATTTATTTGCCAATAGAAACATCAGTGAATTTGCATATATATAC | 1966 |
| | Db | 560 | TWTWAAARAAAWAMAATTAATTTWTTTWTWTWTWTWTWTWTWTAAWMAAWTWTW | 501 |
| | QY | 1967 | TAAATATATATTAACATTTATTTATTTCTTTAATTTTATTTTGTGGCTTTTATTT | 2026 |
| | Db | 500 | AAATTAATAAAAAAAAAAATTTTTTTTTTTTTAAATTAATTAATTTTWTWTWT | 441 |
| | QY | 2027 | ACATTAATTTAACTTGACATATT | 2048 |
| | Db | 440 | TTTWTWTWMAATTAATTTTWT | 419 |
| RESULT 7 | BC036394 | 2672 bp | mRNA | linear |
| LOCUS | DEFINITION | Homo sapiens zinc finger protein 85 (HPF4, HTF1), mRNA (cDNA clone IMAGE5259399), containing frame-shift errors. | | |
| VERSION | ACCESSION | BC036394.1 | GI:23025784 | |
| KEYWORDS | HTC. | | | |
| SOURCE | ORGANISM | Homo sapiens (human) | | |
| REFERENCE | AUTHORS | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (Baes I et al 2002) | | |
| | | Straussberg R.L., Feringold, E.A., Grouse, L.H., Derge, J.G., Klausner R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Bueterov, K.H., Schaefer, C.F., Bat, N.K., Hopkins, R.E., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, P., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stopleon, M., Soares, M.B., Bonaldo, M.P., Casavant, T.L., | | |

| | |
|---------------------------|---|
| TITLE | Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAK Plate: 73 Row: a Column: 16 This clone has the following problem: frame shifted. Location/Qualifiers |
| FEATURES | 1..2672 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:5259399" /tissue_type="Brain, hippocampus" /clone_lib="NIH MGC_95" /lab_host="MDH10B" /note="Vector: pBluescript" |
| ORIGIN | |
| Query Match | 4.5%; Score 92.4; DB 11; Length 2672; |
| Best Local Similarity | 54.1%; Pred. No. 6,1e-05; |
| Matches 236; Conservative | 0; Mismatches 191; Indels 9; Gaps 2 |
| Db | 788 AAAGTATATTGCACATATGAAGGGTGTGATAAAGCCGATATATGACCATCATATTATGA 847 |
| Db | 1745 AGAGAGGCTTACAAATGTTAAAGATGGAAGAAAGCTTTTAAACCAATCCTCAAAACTTAC 180 |
| Db | 848 GCACATTTAAGAACCCACAGTAATGATGCAGCCGATTAAGTACAGTGAAGATTGGA 907 |
| Db | 1805 TGAACATAGAAATTCATCTCTGAGAGAAAGAACCCATGTGATGATG-----GAAAAATGTGG 185 |
| Db | 908 TAAAGCATTTTTCAGAAAATCAATTTGGAAAACACATATTGATACATTCGAAAAAAA 967 |
| Db | 1859 CAAGCTTTTAAACAGTCCTGAATCTTACTAGACATTAAGAAAGATCAATACAGAAAGAGA 191 |
| Db | 968 ACCATTCATTTGTTAGTGTGTGTGAAGGGGTTAATTCGACCAACACTTGAAGAAAGACA 102 |
| Db | 1919 ACCTTACAAATGTGAAGATGTGGCAAAAGTTTAAATGGCCCTCAACCTTATCTATCCA 197 |

QY 1028 TGAATACCCATACAAAGTCATTAAATGACATTGAAATTTGCAAGACATTTTA 1087
| | | | |
DB 1979 TAAGATTAATTCATCGAGAGAAACCATTAATGGAATGTGGCAAACTTTTAA 2038
| | | | |
QY 1088 TAAATCATCATCTTAAACATCATATATATCTGTCTGATGAAAAACATTTACGTG 1144
| | | | |
DB 2039 CCAATCTCAAACTTACCAACATTAAGAAAATTCATACGTGAGAGAAACCTTACATG 2098
| | | | |
QY 1145 TAAACATGTAAATAGTTTCACTGACCTTCACAAATTTAGACACATTAATTTAAACA 1204
| | | | |
DB 2099 TGAAGATGTGGCAAGACCTTTAACCATGCTCTCAACCTTACTTAACATTAAGAAATTC 2158
| | | | |
QY 1205 TCATGTGATCTCT 1220
| | | | |
DB 2159 TACTGGAGAAAACT 2174
| | | | |

RESULT 8
LOCUS A0005136 591 bp DNA linear GSS 27-JUN-1998
DEFINITION CIT-HSP-2290017.TF CIT-HSP Homo sapiens genomic clone 2290017,
genomic survey sequence.
ACCESSION A0005136
VERSION A0005136.1 GI:3082581
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Baas,S., Linher,K.,
AUTHORS Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
Simon,M. and Venter,J.C.
Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)
Unpublished (1998)
Other_GSSs: CIT-HSP-2290017.TR
JOURNAL Contact: Mark Adams
COMMENT Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/cdb/hungen/bac_end_search/bac_end_search.html
Seq primer: M13-21;
Class: BAC ends.
FEATURES
source Location/Qualifiers
1..591
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="GDB:7150844"
/db_xref="taxon:9606"
/clone="2290017"
/sex="Male"
/cell_type="Sperm"
/clone_1b="CIT-HSP"
/note="Vector: pBelobAC11, Site_1: HindIII, Site_2:
HindIII"

ORIGIN
Query Match 4.5%; Score 92; DB 28; Length 591;
Best Local Similarity 52.2%; Pred. No. 0.00017;
Matches 256; Conservative 0; Mismatches 225; Indels 9; Gaps 2;

QY 726 GAAAGTACGAAACCAATGATGATCATCTTAAATCTTCTTCTTCATCAGCTCC 785
| | | | |
DB 69 GAATGGCAAGCCTTCTACCATCTTCAACCTTACACATTAAGGTAATTCAT 128
| | | | |
QY 786 AAAAGTATTTGCAATATGAAGGTGTGATTAAGCCTATTAATGACCATCATTAATTA 845
| | | | |

DB 129 GGAGAGAGCCCTTCAATGTGAGAAATGTGTAAAGCTTTTAAACCACTTCAGCCCTT 188
| | | | |
QY 846 GAGCAACATTTTAAAGAACCCAGATATGATTCGACCGTATTAATGTAACGTGACGATTTG 905
| | | | |
DB 189 ACTACATTAATGATCTTCAATGTTTAAAGAAAACTTACAAATGT-----GAAAGATGT 242
| | | | |
QY 906 GATTAAGCATTTTTCAGAAAATCAATTTGAAACACATATGATATCATCTTCGAAAAA 965
| | | | |
DB 243 GACAAAGCTTTTAAACGATTCATACCTTACTTAACATTAAGTAATATTCATTTGAGAG 302
| | | | |
QY 966 AAACCATTCATTTGTCAGTGTGTGTAAGAGGCTTAATTTCTGACAAACCTTGAAGA 1025
| | | | |
DB 303 AAATCTTACAAATGTGAACAATGTGGCAAGGCTTTAACTGCTTCAACCTTACAAAA 362
| | | | |
QY 1026 CATGAATACCCATCAAGTCATTTAATATGACATTTGAAATTTGTCAAGAAAGCTTT 1085
| | | | |
DB 363 CATAAAAGATTTATCTGAGAGAAACCTTACAAATGTGAAGATGTGGCAAGGCTTT 422
| | | | |
QY 1086 TATTAACAT---CAATCTTAAAGACATCATATATATCTGTGATGAAAAACATTTAACG 1142
| | | | |
DB 423 AATGTCTTTCACACCTTACTACATTAAGATGATTCATACCTGAGAGAAACCTTACAA 482
| | | | |
QY 1143 TGTAAACATGTATATAGTTTCACTGACCTTCAAAATTTAGACACATTAATTTAAA 1202
| | | | |
DB 483 TGTGAAGAAATGTGGCAAGCCTTTAACCACTCTCTCAAACTTACTATACATTAAGATTA 542
| | | | |
QY 1203 CATCATGTG 1212
| | | | |
DB 543 CATACTGGAG 552
| | | | |

RESULT 9
LOCUS B99387 593 bp DNA linear GSS 26-JUN-1998
DEFINITION CIT-HSP-2281A10.TF CIT-HSP Homo sapiens genomic clone 2281A10,
genomic survey sequence.
ACCESSION B99387
VERSION B99387.1 GI:3027197
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Baas,S., Linher,K.,
AUTHORS Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
Simon,M. and Venter,J.C.
Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)
Unpublished (1998)
JOURNAL Contact: Mark Adams
COMMENT Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/cdb/hungen/bac_end_search/bac_end_search.html
Seq primer: M13-21;
Class: BAC ends.
FEATURES
source Location/Qualifiers
1..593
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="GDB:7147044"
/db_xref="taxon:9606"
/clone="2281A10"
/sex="Male"
/cell_type="Sperm"
/clone_1b="CIT-HSP"
/note="Vector: pBelobAC11, Site_1: HindIII, Site_2:
HindIII"

REMARK
COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgapsb-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc.mgc@nih.gov
 Akher, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
 Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
 Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
 Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,
 Maduro, Q.L., Masello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
 McDowell, J., Pearson, R., Stantipod, S., Thomas, P.J., Touchman, J.W.,
 Tsugeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
 Young, A., Zhang, L.-H. and Green, E.D.

FEATURES

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Series: IRAK Plate: 65 Row: a Column: 17
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 13386417
 This clone has the following problem: frame shifted.

Location/Qualifiers
 1..2253

/organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:3674739"
 /tissue_type="Mammary tumor metastasized to lung. Tumor
 arose spontaneously from a senescent normal mammary
 (clonal) outgrowth infected with the virus MMTV."
 /clone_1ib="NCI CGAP_Lu29"
 /lab_host="DHIOB"
 /note="Vector: pCMV-SPORT6"

ORIGIN

Query Match 4.4%; Score 90.4; DB 11; Length 2253;
 Best Local Similarity 51.6%; Pred. No. 0.00013;
 Matches 261; Conservative 0; Mismatches 236; Indels 9; Gaps 2;

710 ACTTATTCATGAGTGAAGTGAACCAATGATATCATCTTATATCTCTTC 769
 810 ACCCTACAAATGTAATCATATGATTAAGCTTTTCAACACAGTTCTTCAAACCTCA 869
 770 TTCTTCATGAGTCCCAAAAGTATATTTGCAATATGAGGGTGTGATTAAGCTTATA 829
 870 TAGAAGAACACATACCGGAGGAAACCTTCAATGTATCAATGATTAAGACCTTTTC 929
 830 TCGACCATCATATTAAGACATTTAGAACCCACATATGATTCGACCGTATTAAG 889
 930 CCGTAAACATTTGTGCAAAACCATATTAAGATCACTCTGAGAAAAACCTTCAAAAG 989
 890 TACAGTGAAGATTTGATTAAGACATTTTTCAGAAAAATCAATTTGGAACACATATGT 949
 990 TA-----ATCATATGATTAAGACCTTTTCAACAGCACTACTCTTCAAACTCAATGAAG 1043
 950 ATCAATTTCCGAAAAAATTCATTTCCATTTCACTGTGTGTAAGGGGTTAATTTCTCG 1009
 1044 AACACATTCGGAGGAAACCTTCAATGTATCAATGATTAAGACCTTTTCTGTATA 1103
 1010 ACAACATTTGAAGACATGAAGATCAACCATCAAAATCATTAATGATTAAGTTGAAA 1069
 1104 ACATAGTCTCCAAACCATATGATTAAGATCACTGAGGAAAAACCTTCAAAATGATCA 1163
 1070 TTGTCAAGAGCATTTTATTAACATCAATCTTTAAGACATCATATATATCTGTTCA 1127
 1164 ATGTATTAAGGCTTTTCAACACATTTCACTCAAAATCATATTAAGACATCACTAG 1223

QY 1128 -GAAAAACATTAAGCTGTGAACATGATATATAAGTTTTCATCGACCTTCAAAATTAAGC 1186
 Db 1224 AGAGATGCTCTCAAAATGATATGATATGATCAACAAAGCTTTTATCATTAAGAAATCTTCA 1283
 QY 1187 ACAACATTAATTAACATCAATGATG 1212
 Db 1284 GATTCATTAAGTAAACATCACTGAG 1309

RESULT 12
 LOCUS BC022527
 DEFINITION Homo sapiens hypothetical protein LOC148206, mRNA (cDNA clone
 IMAGE:479757), containing frame-shift errors.
 ACCESSION BC022527
 VERSION BC022527.1 GI:18490307
 KEYWORDS HTC.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Scheefer, C.F., Bhat, N.K.,
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
 Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
 Stappleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
 Schetz, T.E., Brownstein, W.J., Usdin, T.B., Toshiyuki, S.,
 Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
 Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.D.,
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
 Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hultk, S.W.,
 Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
 Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,
 Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,
 Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
 Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Small, D.E.,
 Scherch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 TITLES
 JOURNAL

REMARK
 COMMENT
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgapsb-remail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
 Toshiyuki and Piero Carninci (RIKEN)
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome
 Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) mcdpax11.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
 R. M.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Series: IRAK Plate: 32 Row: f Column: 2
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Hexamer frequency ORF
 analysis, similarity but not identity to protein
 This clone has the following problem: frame shifted.

| | | | |
|----|------|--|------|
| QY | 846 | GAGCAACATTTAAGAACCAGTAATGATGACCGTAAATGTAACAGTGGACGATTTG | 905 |
| Db | 1162 | ACTACACATAAGTTCATTCATGTTAAAGAAAAAACCCTACCAATG-----GAAGAAATGT | 1215 |
| QY | 906 | GATAAAGCATTTTCAGAAATTCACATTTGGAAACACATATTGTATCATTTCCGAAAA | 965 |
| Db | 1216 | GACAAAGCTTTTAACCGATTTCTCAATCTTACTAAACCTAAGATTAATTCATTTCTGGAGAG | 1275 |
| QY | 966 | AAACCATTCATTTGTCAGTGTGTGTAAAGGGGTTAATTCTGACAAACATTTGAAAAA | 1025 |
| Db | 1276 | AAATCTTCAAAATGTGAACAATGTGGCAAAAGCTTTAACTGCTCAACCTTACAAAA | 1335 |
| QY | 1026 | CATGAATACACCACATCAAAATCATTTAAATGTACATTTGAAAATGTGCAGGAAGCATTT | 1085 |
| Db | 1336 | CATGAACAATTTCAATCTGAGAGAAAAACCTTACAAATGTGAAGAATGTGGCAAAAGCCTTT | 1395 |
| QY | 1086 | TATTAACAAT---CATCTTTAAGACATCATATTATTCGTGTCATGAAAAAACATTAACG | 1142 |
| Db | 1396 | AATGTGCTTACAGCCTTACTACACATTAAGATGATTCATCTGGAGAGAAACCTTACAA | 1455 |
| QY | 1143 | TGTAAACATGTAATAAAGTTTCACTCGACCTTCAAAATTAGACAAACATTAATTAAAA | 1202 |
| Db | 1456 | TGTGAAGATGTGGCAAAAGCCTTTAACCACATCTCCAAACCTTACTATACATAAGTATTT | 1515 |
| QY | 1203 | CATCATGGTGAATCTCCT | 1220 |
| Db | 1516 | CATCTGGAGAAAACT | 1533 |

| RESULT 14 | | | |
|------------|--|------------|-------------|
| LOCUS | BE161630 | 604 bp | mRNA linear |
| DEFINITION | MR-HT0446-260300-201-f06 HT0446 Homo sapiens CDNA, mRNA sequence. | | |
| ACCESSION | BE161630 | | |
| VERSION | BE161630.1 | GI:8624351 | |
| KEYWORDS | EST. | | |
| SOURCE | Homo sapiens (human) | | |
| ORGANISM | Homo sapiens | | |
| | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| | Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | |
| REFERENCE | 1 (bases 1 to 604) | | |
| AUTHORS | Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R., | | |

| | |
|----------|---|
| TITLE | Shotgun sequencing of the human transcriptome with ORF expressed sequence tags |
| JOURNAL | Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000) |
| MEDLINE | 20202863 |
| PMID | 10737800 |
| COMMENT | Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/ICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=kt2-MR3-HR0446-260300-201-f06&kt=2000-03-26&kt4=1) Seq primer: puc.18 forward High quality sequence stop: 583. Location/Qualifiers 1..604 |
| FEATURES | |
| source | |

```
FEATURES
source
location/Qualifiers
1..604
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_id="HT0446"
```

/note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
 Site_2: SmaI; A mini-library was made by cloning products
 derived from ORSTS PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the pUC 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."

| Query Match | 4.3%; | Score 89.2; | DB 10; | Length 604; |
|-----------------------|-----------------|--|-----------|-------------|
| Best Local Similarity | 53.7%; | Pred. No. 0.00044; | | |
| Matches 234; | Conservative 0; | Mismatches 133; | Indels 9; | Gaps 2; |
| Qy | 788 | AAAGTATATTTGCACATATGAAAGGTGTGATAAAGCCTATATATGACCATCATATATATAGA | 847 | |
| | | | | |
| Db | 102 | AGAGAAACCTTACCAATGTGAMAGATGTGACAAAGTTTTTAAAGTGTGCTAATACACACCTTGA | 161 | |
| Qy | 848 | GCAACATTTTAAAGAACCCACAGTAATGATGACCGGTATAAATGTACAGTGACATTTGGA | 907 | |
| Db | 162 | AAGGCAATAGGAGAAATTCATCTGAGAGAGAAACCTTACCAATGTAAAGT-----TTGTGA | 215 | |
| Qy | 908 | TAAAGCATTTTTCAGAAAATTCACATTGGAAAACATATTTGTATCAGATTCGAGAAAAAA | 967 | |
| Db | 216 | CAAGGCTTTTACAGAGTGAATTCACACCTGGACACAAATATGTAAATTCACATCGAGAGAA | 275 | |
| Qy | 968 | ACCAATTCATTTGTTCACTGTGTGTGTAAGGGGTTAATTTTCGACAAACCTGAAAAGCA | 1027 | |
| Db | 276 | ACCTTACAAAGTATATAGTGTGTGACAGACCTTTGTTCAAAATTCATCTCTTTGTAATGCA | 335 | |
| Qy | 1028 | TGAATATCACCCATG--AAAGTCATTTAAATGTACATTTGAAAAATTTGTCAAGAGCAATT | 1084 | |
| Db | 336 | TAAAGTCATTCATACACTGAGAGAAATGTTACAGGTGTAAGAATGTGGCAAGGTTTTTAA | 395 | |
| Qy | 1085 | TTATTAACATCAATCTTTAAGACATCATATATATATCTGTTTCATGAAAAAAACATTAACGTG | 1144 | |
| Db | 396 | TCACAAATCAAACTTGTCATGTCAATCTATAGACTTCATCTGAGAGAGAAACCTTACAAAGTG | 455 | |
| Qy | 1145 | TAAACAAATGTAATTAAGTTTTCACCTGACCTTCAAAATTTGACACACATTAATTTAAACA | 1204 | |
| Db | 456 | TAAATGATGTGGCAGAGGTTTTTAATGAAAAATCAAACTTGAAACATCATATAGATTTCA | 515 | |
| Qy | 1205 | TCATGTGTGATCTGCT | 1220 | |
| Db | 516 | TATCGAGAAAAACCT | 531 | |

| | |
|------------|---|
| RESULT | 15 |
| BC032590 | |
| LOCUS | |
| DEFINITION | BC032590 Homo sapiens cDNA clone IMAGE:5502691, containing frame-shift errors. |
| ACCESSION | BC032590 |
| VERSION | BC032590.1 GI:2161971 |
| KEYWORDS | HNC. |
| SOURCE | Homo sapiens (human) |
| ORGANISM | Homo sapiens |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 3609) Straussberg R.L., Peingold,E.A., Grouse,L.H., Derye,J.G., Klauener,R.D., Collins,F.S., Wagner,I., Sherman,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Bietlow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,Y., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.M., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldi,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.U., Uddin,T.B., Toshiyuki,S., Cherniack,P., Prange,C., Kaha,S.S., Loquellano,N.A., Peters,G.U., Abramson,R.D., Mullaly,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulik,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodriguez,S., Sanchez,A., Whitting,M., Madan,A., Young,A.C., Shevchenko,Y, |

Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
 Butlerfield, Y.S., Krzywinski, M.I., Skalska, U., Smalins, D.E.,
 Schnerch, A., Schein, J.E., Jones, S.J., and Mair, M.A.
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

REMARK
 COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNI)
 DNA Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc_mgc@hgrl.nih.gov
 Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
 Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
 Dietrich, N.L., Granite, S., Guan, J., Gupta, P., Haghighi, P.,
 Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Latic, P., Legaspi, R.,
 Maduro, O.L., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
 McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,
 Tsurgoun, C., Vogt, U.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
 Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/ILNI at: <http://image.llnl.gov>
 Series: IRAK Plate: 69 Row: g Column: 20
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 13430873
 This clone has the following problem: frame shifted.

FEATURES
 source

1..3609
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5502691"
 /tissue_type="Eye, retinoblastoma"
 /clone_id="NIH_MGC_67"
 /lab_host="DH10B"
 /note="Vector: PCMV-SPORT6"

ORIGIN

Query Match 4.3%; Score 89.2; DB 11; Length 3609;
 Best Local Similarity 53.7%; Pred. No. 0.00015;
 Matches 234; Conservative 0; Mismatches 193; Indels 9; Gaps 2;

Db 788 AAAGTATATTGGCACTATGAGGGGTGATTAAGCCTATATGACCATCATTTATTGA 847
 587 AGAGAAACCTTACAAATGTGAAGATGTGACAAAGTTTGTGCTAATACACACCTTGA 646
 Qy 848 GCACATTTTGAACCCACAGTATGATGACCGGTAAATGTACAGTGAAGCATTTGA 907
 Db 647 AAGCATATGAGAAATTTCTACTGAGAGAAACCGTACAAATGTAAAGT-----TTGTGA 700
 Qy 908 TAAAGCATTTTTCAGAAATTCATTTGGAACACATATGTATCATATTCGAAAAA 967
 Db 701 CAAAGCTTTTCAGAGTATTCACACCTGGACAAACATATTTGTAATTCACACTGAGAGAA 760
 Qy 968 ACCATTCCATTGTCAGTGTGTGTGTAAGGGGTTAATTTCTGACACACTGAAAAACA 1027
 Db 761 ACCTTACAAGTATATGATGTGTGCAAGACCTTGTTCAAAATTCATCTCTTGAATGCA 820

Qy 1028 TGAATACCCATAC--AAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAGATT 1084
 Db 821 TAAGTCATTTCAATCTGAGAGAAATGTTCAAAGTAAATGAATGTGCAAGGTTTAA 880
 Qy 1085 TTAATTAACATCAATCTTTAAGACATCAATTAATTTCTGTTCAATGAAAAACATTAACGTG 1144
 Db 881 TCAGAAATCAACCTTGCAATGTATCATATAGACTTCATACAGAGAGAAACCTTACAAAGTG 940
 Qy 1145 TAAACATTTAATTAAGTTTTCACCTGACCTTCAAAAATTAGACACATTAATTTAAAAACA 1204
 Db 941 TAAATGAATGTGCGAAGGTTTAAATGAAAAATCAACCTTGACATCATATAGAGTTCA 1000
 Qy 1205 TCATGTGATTCCT 1220
 Db 1001 TATCGAGAAAAAACCCT 1016

Search completed: July 28, 2004, 18:49:46
 Job time : 5418 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2004, 16:48:02 ; Search time 54 Seconds

(without alignments)
2155.733 Million cell updates/sec

Title: US-09-831-804-3

Perfect score: 412

Sequence: 1 MSESDETKSISLIISSSSSS.....PIWKARMDLPNETSVISR 412

Scoring table: OLIGO

Searched: 1586107 seqs, 282547505 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

1: geneseqp19808.*
2: geneseqp19908.*
3: geneseqp20008.*
4: geneseqp20018.*
5: geneseqp20028.*
6: geneseqp20038.*
7: geneseqp20048.*
8: geneseqp20058.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1 | 412 | 100.0 | 412 | 3 | AA93316 |
| 2 | 352 | 85.4 | 412 | 5 | ABP73656 |
| 3 | 10 | 2.4 | 546 | 5 | ABBS7056 |
| 4 | 9 | 2.2 | 58 | 4 | ABG19930 |
| 5 | 9 | 2.2 | 114 | 4 | AAU30108 |
| 6 | 9 | 2.2 | 114 | 4 | AAU30387 |
| 7 | 9 | 2.2 | 128 | 3 | AA857179 |
| 8 | 9 | 2.2 | 168 | 5 | ABP64928 |
| 9 | 9 | 2.2 | 176 | 6 | ADA55380 |
| 10 | 9 | 2.2 | 202 | 5 | ABP28933 |
| 11 | 9 | 2.2 | 214 | 7 | ADD47645 |
| 12 | 9 | 2.2 | 214 | 7 | ADR60447 |
| 13 | 9 | 2.2 | 214 | 7 | ADRS7980 |
| 14 | 9 | 2.2 | 214 | 7 | ADRS7984 |
| 15 | 9 | 2.2 | 214 | 7 | ADRS7984 |
| 16 | 9 | 2.2 | 214 | 7 | ADRS7984 |
| 17 | 9 | 2.2 | 215 | 6 | ABU07428 |
| 18 | 9 | 2.2 | 215 | 6 | ABU07428 |
| 19 | 9 | 2.2 | 215 | 6 | ABU07428 |
| 20 | 9 | 2.2 | 215 | 7 | ADD40789 |
| 21 | 9 | 2.2 | 215 | 7 | ADD40789 |
| 22 | 9 | 2.2 | 216 | 6 | AA835866 |
| 23 | 9 | 2.2 | 220 | 4 | ABG18955 |
| 24 | 9 | 2.2 | 223 | 4 | AAU30109 |
| 25 | 9 | 2.2 | 226 | 4 | ABH12356 |
| | | | 229 | 5 | ABP64829 |

| | | | | | |
|----|---|-----|-----|---|----------|
| 26 | 9 | 2.2 | 233 | 4 | ABG19932 |
| 27 | 9 | 2.2 | 237 | 4 | ABG09807 |
| 28 | 9 | 2.2 | 240 | 4 | AAU30388 |
| 29 | 9 | 2.2 | 366 | 3 | AA838042 |
| 30 | 9 | 2.2 | 372 | 3 | AA832376 |
| 31 | 9 | 2.2 | 571 | 6 | ABRS2911 |
| 32 | 9 | 2.2 | 589 | 7 | ABBS5263 |
| 33 | 9 | 2.2 | 699 | 7 | AAV67598 |
| 34 | 9 | 2.2 | 830 | 5 | ABR04808 |
| 35 | 9 | 2.2 | 830 | 5 | ABR04815 |
| 36 | 8 | 1.9 | 13 | 3 | ABR01687 |
| 37 | 8 | 1.9 | 23 | 6 | ABO12946 |
| 38 | 8 | 1.9 | 23 | 6 | ABO13305 |
| 39 | 8 | 1.9 | 29 | 4 | AAW2536 |
| 40 | 8 | 1.9 | 29 | 4 | AAW2277 |
| 41 | 8 | 1.9 | 29 | 4 | AAW2277 |
| 42 | 8 | 1.9 | 29 | 5 | ABG42092 |
| 43 | 8 | 1.9 | 63 | 4 | AAW85534 |
| 44 | 8 | 1.9 | 96 | 4 | ABG27789 |
| 45 | 8 | 1.9 | 115 | 7 | ADE58420 |

ALIGNMENTS

RESULT 1
ID AA93316 standard; protein, 412 AA.

AA93316;

04-SEP-2000 (first entry)

A transcription factor designated CATPITTA.

Transcription factor; CATPITTA; DNA-binding protein;
ribosomal RNA 5S gene; fungal infection.

Candida albicans.

Key Location/Qualifiers

Misc-difference 193 /note= "Ser encoded by CTG"

Misc-difference 339 /note= "Ser encoded by CTG"

WO200028037-A1.

18-MAY-2000.

09-NOV-1999; 99WO-FR002739.

10-NOV-1998; 98FR-00014147.

(HMRI) HOECHST MARION ROUSSEL.

Bordon-Pallier F, Camier S, Senenac A;

WPI, 2000-376549/32.

N-PSDB; AAA15398.

New nucleic acid encoding Candida albicans transcription factor, useful
e.g. in screening for antitumor agents and for immunization.

Claim 12; Page 35-36; 45pp; French.

The present sequence represents a Candida albicans transcription factor,
designated CATPITTA. The polypeptide is a DNA-binding protein, which is
involved in initiating transcription of the ribosomal RNA 5S gene. The
polynucleotide is used to screen for its specific inhibitors, potentially
useful as antitumor agents, to raise an antibody response that is
protective against fungal infection and to raise antibodies. Such
antibodies, as well as the polypeptides and polynucleotides are used in

CC compositions for diagnosing and treating fungal infections, e.g. by
 CC detecting polymorphisms and mutations

XX Sequence 412 AA;

Query Match 100.0%; Score 412; DB 3; Length 412;
 Best Local Similarity 100.0%; Pred. No.0;
 Matches 412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSESDETKSISLSSSSSRPKYICTYEGCDKAVNRPSELBOHLRTSHNDRPYKCTVD 60
 DB 1 MSESDETKSISLSSSSSRPKYICTYEGCDKAVNRPSELBOHLRTSHNDRPYKCTVD 60
 QY 61 DCKAFPRKSHLETHIVSHSEKRPFHGCVGKGVNSRQHLKRHEITHTKSPKCTFENCQ 120
 DB 61 DCKAFPRKSHLETHIVSHSEKRPFHGCVGKGVNSRQHLKRHEITHTKSPKCTFENCQ 120
 QY 121 AFYKQSLRHILISVHEKTLTKQCKNVFTRPSKLAQHLKHHGSPAYQCDHGGCTKNF 180
 DB 121 AFYKQSLRHILISVHEKTLTKQCKNVFTRPSKLAQHLKHHGSPAYQCDHGGCTKNF 180
 QY 181 QTWSVLQFHIKQSHPKLCKPCGKGVGKGLSSHMLSHDSTMIKIMTCDYCDVGKFAK 240
 DB 181 QTWSVLQFHIKQSHPKLCKPCGKGVGKGLSSHMLSHDSTMIKIMTCDYCDVGKFAK 240
 QY 241 KNEIYEHYNI FHDGNITPDDLKETEYVKLENLDQGSKLNHLETEKLVKVEDDEDEE 300
 DB 241 KNEIYEHYNI FHDGNITPDDLKETEYVKLENLDQGSKLNHLETEKLVKVEDDEDEE 300
 QY 301 DSIDEKSDVRSMSAQRSTKSTFTASLEGSYSKILNSGKKINCPCXNCDMFSEKREY 360
 DB 301 DSIDEKSDVRSMSAQRSTKSTFTASLEGSYSKILNSGKKINCPCXNCDMFSEKREY 360
 QY 361 DLRRLKMHNDNLQRIEFLNSIEKEETPEGEPLVYKARMDLLPNETSVISR 412
 DB 361 DLRRLKMHNDNLQRIEFLNSIEKEETPEGEPLVYKARMDLLPNETSVISR 412

RESULT 2
 ABP73656
 ID ABP73656 standard; protein; 412 AA.

XX AC ABP73656;

XX 30-JAN-2003 (first entry)

XX Candida albicans essential protein SEQ ID NO 7493.

XX Fungus: yeast; tetracycline; promoter; GRACE strain; biosynthesis;
 KM signal transduction; DNA replication; cell division; growth;
 KM proliferation; Candida albicans; fungicide; antifungal.

XX Candida albicans.

XX WO200253728-A2.

XX 11-JUL-2002.

XX 26-DEC-2001; 2001WO-US049486.

XX 29-DEC-2000; 2000US-0259128P.

XX 20-FEB-2001; 2001US-00792024.

XX 22-AUG-2001; 2001US-0314050P.

XX (ELIT-) ELITRA PHARM INC.

XX Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;

XX WPI; 2002-566694/60.

XX N-PSDB; ABZ32206.
 XX Constructing strains for identifying gene products as effective targets
 PT for therapeutic intervention, by inactivating in the strain one allele of

PT a gene and placing other allele of the gene under conditional expression.
 XX Claim 44; SEQ ID NO 7493; 167pp + Sequence Listing; English.

XX The invention relates to constructing (M1) a strain of diploid fungal
 CC cells in which both alleles of a gene are modified, comprising modifying
 CC one allele by insertion or replacement by a cassette having an
 CC expressible selectable marker and modifying other allele by
 CC recombination, of a promoter replacement fragment with a heterologous
 CC promoter, so that expression of the second allele is regulated by the
 CC promoter. (M1) is useful for constructing a strain of diploid fungal
 CC cells in which both alleles of a gene are modified. The diploid fungal
 CC cells having both alleles modified are useful for identifying a gene that
 CC is essential to the survival or growth of a fungus, a gene that
 CC contributes to the virulence and/or pathogenicity of a fungus, a gene
 CC that contributes to the resistance of a diploid fungus to an antifungal
 CC agent, an antifungal agent that inhibits the growth of a diploid fungus
 CC and for identifying a therapeutic agent for treatment of a mammalian
 CC disease. (M1) is useful for identifying a compound which modulates the
 CC activity of a gene product, preferably enzymatic activity, carbon
 CC compound catabolism, biosynthetic, transporter, transcriptional,
 CC translational, signal transduction, DNA replication and cell division
 CC activity. The method is useful for identifying a compound having the
 CC ability to inhibit growth or proliferation of C. albicans cells and for
 CC treating infection by C. albicans. The present sequence is that of an
 CC essential Candida albicans protein used in the method of the invention.
 CC Note: The sequence data for this patent is not represented in the printed
 CC specification but is based on sequence information supplied to Derwent by
 CC the European Patent Office

XX Sequence 412 AA;

Query Match 85.4%; Score 352; DB 5; Length 412;
 Best Local Similarity 100.0%; Pred. No.0;
 Matches 352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 DCKAFPRKSHLETHIVSHSEKRPFHGCVGKGVNSRQHLKRHEITHTKSPKCTFENCQ 120
 DB 61 DCKAFPRKSHLETHIVSHSEKRPFHGCVGKGVNSRQHLKRHEITHTKSPKCTFENCQ 120
 QY 121 AFYKQSLRHILISVHEKTLTKQCKNVFTRPSKLAQHLKHHGSPAYQCDHGGCTKNF 180
 DB 121 AFYKQSLRHILISVHEKTLTKQCKNVFTRPSKLAQHLKHHGSPAYQCDHGGCTKNF 180
 QY 181 QTWSVLQFHIKQSHPKLCKPCGKGVGKGLSSHMLSHDSTMIKIMTCDYCDVGKFAK 240
 DB 181 QTWSVLQFHIKQSHPKLCKPCGKGVGKGLSSHMLSHDSTMIKIMTCDYCDVGKFAK 240
 QY 241 KNEIYEHYNI FHDGNITPDDLKETEYVKLENLDQGSKLNHLETEKLVKVEDDEDEE 300
 DB 241 KNEIYEHYNI FHDGNITPDDLKETEYVKLENLDQGSKLNHLETEKLVKVEDDEDEE 300
 QY 301 DSIDEKSDVRSMSAQRSTKSTFTASLEGSYSKILNSGKKINCPCXNCDMFSEKREY 360
 DB 301 DSIDEKSDVRSMSAQRSTKSTFTASLEGSYSKILNSGKKINCPCXNCDMFSEKREY 360
 QY 361 DLRRLKMHNDNLQRIEFLNSIEKEETPEGEPLVYKARMDLLPNETSVISR 412
 DB 361 DLRRLKMHNDNLQRIEFLNSIEKEETPEGEPLVYKARMDLLPNETSVISR 412

RESULT 3
 ABB57056
 ID ABB57056 standard; protein; 546 AA.

XX AC ABB57056;

XX 07-MAR-2002 (first entry)

XX Mouse ischaemic condition related protein sequence SEQ ID NO:106.

XX Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;
 KM vasospastic ischaemia; ischaemic condition; ischaemic disease.

XX OS Mus musculus.
XX XX WO200188188-A2.
XX PN 22-NOV-2001.
XX PD 18-MAY-2001; 2001WO-JP004192.
XX PF 18-MAY-2000; 2000JP-00145977.
XX PR 18-MAY-2000; 2000JP-00145977.
XX PA (UYNI-) UNIV NIHON SCHOOL JURIDICAL PERSON.
XX PI Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;
XX DR WPI; 2002-034733/04.
XX DR N-PSDB; ABI99270.
XX PT Examining the ischemic condition (e.g. occlusive ischemia) by measuring
PT expression levels of particular genes defined in the specification or by
PT determining the expression profile of a gene group comprising these
PT genes.
XX PS Claim 2; Page 311-313; 2690pp; English.
XX CC The present invention describes a method for examining ischaemic
CC conditions, comprising measuring the expression levels of particular
CC genes (I) in a test sample or determining the expression profile of a
CC gene group in the sample comprising genes selected from (I). The method
CC is useful for examining the ischaemic condition (e.g. compressive
CC ischemia, occlusive ischaemia or vasospastic ischaemia) by measuring the
CC expression levels of particular genes (ABI99202 to ABI99912, encoding the
CC protein sequences in ABB57020 to ABB57374) or by determining the
CC expression profile of a gene group comprising these genes. The expression
CC levels or expression profiles produced by these genes are used as an
CC indicator when screening for ischaemic condition-improving drugs or
CC therapeutics for ischaemic diseases. ABI99913 and ABI99914 represent PCR
CC primers for a mouse ischaemic condition related sequence, which are used
CC in the exemplification of the present invention
XX SQ Sequence 546 AA;
XX
XX Query Match 2.4%; Score 10; DB 5; Length 546;
XX Best Local Similarity 100.0%; Pred. No. 1.4;
XX Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 100 LKRRHITHTK 109
DB 505 LKRRHITHTK 514
RESULT 4
ABG19930
ID ABG19930 standard; protein; 58 AA.
XX AC ABG19930;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #19921.
XX KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KM Food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX PR 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
XX PA Dmanac RT, Liu C, Tang YT;
XX PI WPI; 2001-639362/73.
XX DR N-PSDB; AAS84117.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX PS Claim 20; SEQ ID NO 50289; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABB00010-ABB30377 represent novel human diagnostic
CC patent did not appear in the printed specification. Note: The sequence data for this
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 58 AA;
XX
XX Query Match 2.2%; Score 9; DB 4; Length 58;
XX Best Local Similarity 100.0%; Pred. No. 1.7;
XX Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 293 EEDDEDEED 301
DB 31 EEDDEDEED 39
RESULT 5
AAU30108
ID AAU30108 standard; protein; 114 AA.
XX AC AAU30108;
XX DT 18-DEC-2001 (first entry)
XX DE Novel human secreted protein #599.
XX KM Human; vaccination; gene therapy; nutritional supplement;
XX KM stem cell proliferation; haematopoiesis; nerve tissue regeneration;
XX KM immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX OS Homo sapiens.
XX PN WO200179449-A2.
XX PD 25-OCT-2001.
XX PF 16-APR-2001; 2001WO-US008656.
XX PR 18-APR-2000; 2000US-00552929.
XX PR 26-JAN-2001; 2001US-00770160.

XX (HXSE-) HXSEQ INC.
PA
XX
XX Tang YT, Liu C, Dymnmac RT;
PI
XX
XX WPI, 2001-611725/70.
DR
XX
XX
PT Nucleic acids encoding a range of human polypeptides, useful in genetic
PT vaccination, testing and therapy.
XX
XX
XX
PS Claim 20; Page 241; 765bp; English.
XX
XX The invention relates to novel human secreted polypeptides. The
CC polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising the
CC nucleic acids encoding the polypeptides and cells genetically engineered
CC to express them are also useful for producing the proteins. The proteins
CC are useful in genetic vaccination, testing and therapy, and can be used
CC as nutritional supplements. They may be used to increase stem cell
CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
CC and/or nerve tissue growth or regeneration; immune suppression and/or
CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
CC AADZ9510-AADZ3304 represent the amino acid sequences of novel human
CC secreted proteins of the invention
XX
XX Sequence 114 AA;
SQ

| | Query Match | Similarity | 2.2%: Score 9; | DB 4; | Length 114; |
|------------|---|----------------------------|----------------|----------------|-------------|
| Best Local | 9; | Conservative | 100.0%; | Pred. No. 3.1; | |
| Matches | 9; | Conservative | 0; | Mismatches | 0; |
| | | | | Indels | 0; |
| | | | | Gaps | 0; |
| Oy | 293 | EDEDDEED | 301 | | |
| | | | | | |
| | | | | | |
| | | | | | |
| Db | 86 | EDEDDEED | 94 | | |
| RESULT 6 | | | | | |
| AAU30387 | | | | | |
| ID | AAU30387 | standard; protein; 114 AA. | | | |
| XX | AAU30387; | | | | |
| XX | 18-DEC-2001 | (first entry) | | | |
| XX | Novel human secreted protein #878. | | | | |
| XX | Human; vaccination; gene therapy; nutritional supplement; | | | | |
| KM | stem cell proliferation; haematopoiesis; nerve tissue regeneration; | | | | |
| KW | immune suppression; immune stimulation; anti-inflammatory; leukaemia. | | | | |
| XX | Homo sapiens. | | | | |
| OS | WO200179449-A2. | | | | |
| XX | 25-OCT-2001. | | | | |
| XX | 16-APR-2001; 2001WO-US008656. | | | | |
| XX | 18-APR-2000; 2000US-00552929. | | | | |
| PR | 26-JAN-2001; 2001US-00770160. | | | | |
| XX | (HYSB-) HYSBQ INC. | | | | |
| PA | Tang YT, Liu C, Drmanac RT; | | | | |
| XX | WPI, 2001-611725/70. | | | | |
| XX | Nucleic acids encoding a range of human polypeptides, useful in genetic | | | | |

| | |
|----|--|
| EP | vaccination, testing and therapy. |
| XX | |
| PS | Claim 20; Page 287; 765pp; English. |
| XX | |
| CC | The invention relates to novel human secreted polypeptides. The |
| CC | polypeptides and antibodies to the polypeptides are useful for |
| CC | determining the presence of or predisposition to a disease associated |
| CC | with altered levels of polypeptide. The polypeptides are also useful for |
| CC | identifying agents (agonists and antagonists) that bind to them. Cells |
| CC | expressing the proteins are useful for identifying a therapeutic agent |
| CC | for use in treatment of a pathology related to aberrant expression or |
| CC | physiological interactions of the polypeptide. Vectors comprising the |
| CC | nucleic acids encoding the polypeptides and cells genetically engineered |
| CC | to express them are also useful for producing the proteins. The proteins |
| CC | are useful in genetic vaccination, testing and therapy, and can be used |
| CC | as nutritional supplements. They may be used to increase stem cell |
| CC | proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon |
| CC | and/or nerve tissue growth or regeneration; immune suppression and/or |
| CC | stimulation; as anti-inflammatory agents; and in treatment of leukaemias. |
| CC | AAD29510-AU23304 represent the amino acid sequences of novel human |
| CC | secreted proteins of the invention |
| XX | |
| SQ | Sequence 114 AA; |
| | |
| | Query Match 2.2%; Score 9; DB 4; Length 114; |
| | Best Local Similarity 100.0%; Pred. No. 3.1; |
| | Matches 9; Conservative 0; Mismatches 0; Indels 0; |
| QY | 293 EDEDEDDEP 301 |
| | |
| DB | 86 EDEDEDDEED 94 |

| | |
|----------|---|
| RESULT 7 | |
| AAB57179 | |
| ID | AAB57179 standard; protein, 128 AA. |
| XX | |
| AC | AAB57179; |
| XX | |
| DT | 13-MAR-2001 (first entry) |
| XX | |
| DE | Human prostate cancer antigen protein sequence SEQ ID NO:1757. |
| XX | |
| KW | Human; prostate cancer; prostate cancer antigen; detection; diagnosis; |
| KW | neuroprotective; cytosolic; cardioactive; immunomodulatory; muscular; |
| KW | vulnerary; gastrointestinal; nephrotoxic; antiinfective; gynaecological; |
| KW | antibacterial; gene therapy; neural; immune; reproductive; renal; |
| KW | gastrointestinal; pulmonary; cardiovascular; proliferative disorder; |
| KW | wound; infectious disease. |
| XX | |
| OS | Homo sapiens. |
| XX | |
| PN | WO20055174-A1. |
| XX | |
| PD | 21-SEP-2000. |
| XX | |
| PF | 08-MAR-2000; 2000WO-US005988. |
| XX | |
| PR | 12-MAR-1999; 99US-0124270P. |
| XX | |
| PA | (HUMA-) HUMAN GENOME SCI INC. |
| PA | (ROSE/) ROSEN C A. |
| XX | |
| PI | Rosen CA, Ruben SM; |
| XX | |
| DR | WPI, 2000-587513/55. |
| DR | N-PSDB; AAF16382. |
| XX | |
| PT | Prostate cancer associated gene sequences, referred to as prostate cancer |
| PT | antigens, useful for treatment, prevention, and diagnosis of disorders |
| PT | such as prostate cancer. |
| XX | |
| PS | Claim 11; Page 2225-2226; 2338pp; English. |

XX AAF15566 to AAF16505 encode the human prostate cancer associated
CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC The prostate cancer antigens can have neuroprotective, cytostatic,
CC cardioactive, immunomodulatory, muscular, vulnerrary, gastrointestinal,
CC nephrotropic, antiinfective, gynaecological and antibacterial activities,
CC and can be used in gene therapy. The prostate cancer antigen
CC polynucleotides may be used for detection of prostate cancer, chromosome
CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The prostate cancer antigens may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders, wounds, and infectious diseases. AAF15506 to AAF16514 to
CC AAB57303 represent sequences used in the exemplification of the present
CC invention
CC
SQ Sequence 128 AA;
Query Match 2.2%; Score 9; DB 3; Length 128;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 293 EDEDEDEED 301
Db 105 EDEDEDEED 113
RESULT 8
ABP64928
ID ABP64928 standard; protein; 168 AA.
XX
XX ABP64928;
XX
DT 25-FEB-2003 (first entry)
XX
XX Human protein SEQ ID 588.
DE
XX
XX Human; expressed sequence tag; EST; haematopoietic disorder;
KW central nervous system disease; viral infection;
KW peripheral nervous system disease; non-healing wound; infectious disease;
KW immune deficiency; immune disorder; bacterial infection; allergy/ cancer;
KW fungal infection; autoimmune disorder; coagulation disorder; nootropic;
KW antiallergic; antiinflammatory; immunosuppressive; neuroprotective;
KW cytostatic; haemostatic; virucide; antibacterial; fungicide;
KW immunostimulant; cerebroprotective.
XX
OS Homo sapiens.
XX
XX
PN WO200259260-A2.
XX
PD 01-AUG-2002.
XX
XX 16-NOV-2001; 2001WO-US042950.
PF
XX 17-NOV-2000; 2000US-00714936.
PR
XX (HYSB-) HYSBQ INC.
XX
XX Tang YT, Goodrich RM, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA;
PI Ren F, Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX WPI; 2002-590824/63.
DR N-PSDB; ABQ99514.
XX
XX New isolated polynucleotide, useful in research, diagnostic or
PT therapeutic methods, e.g. preventing or treating disorders involving
PT aberrant protein expression or biological activity.
XX
XX Claim 20; SEQ ID NO 588; 394pp; English.
PS
XX The present invention relates to novel human coding sequences (ABQ99268-
CC ABQ99608) and proteins (ABP64682-ABP65022). The sequences are useful in
CC therapeutic, diagnostic and research methods. The polynucleotides may be

CC used in the field of molecular biology as hybridisation probes, primers
CC for PCR, for chromosome and gene mapping, for the recombinant production
CC of protein, or in generation of anti-gene DNA or RNA. The
CC polynucleotides are useful in diagnostics as expressed sequence tags
CC (ESTs) for identifying expressed genes or for physical mapping of the
CC human genome. The proteins may be used as molecular weight markers, or as
CC nutritional sources or supplements. The proteins may be used to maintain
CC and expand cell population in a totipotent or pluripotent state
CC useful for re-engineering damaged or diseased tissues, transplantation,
CC manufacture of bio-pharmaceuticals or the development of bio-sensors. The
CC polynucleotides and proteins are useful for preventing, treating or
CC ameliorating disorders involving aberrant protein expression or
CC biological activity, e.g. haematopoietic disorders, central/peripheral
CC nervous system diseases, mechanical and traumatic disorders, non-healing
CC wounds, immune deficiencies and disorders, infectious diseases caused by
CC viral, bacterial or fungal infection, autoimmune disorders, allergic
CC reactions and conditions, coagulation disorders, or cancer. The
CC polynucleotide sequences of the invention were assembled from ESTs
CC isolated mainly by sequencing by hybridisation, and in some cases,
CC sequences obtained from one or more public databases. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
CC
SQ Sequence 168 AA;
Query Match 2.2%; Score 9; DB 5; Length 168;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 293 EDEDEDEED 301
Db 142 EDEDEDEED 150
RESULT 9
ADA55380
ID ADA55380 standard; protein; 176 AA.
XX
XX ADA55380;
XX
AC ADA55380;
XX
DT 20-NOV-2003 (first entry)
XX
XX Human protein, SEQ ID 2948.
DE
XX
XX Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KW Gene Therapy; human; secretory protein; membrane proteins; cancer;
KW inflammatory disease; osteoporosis; neurological disease.
XX
XX Homo sapiens.
XX
XX
PN EP1293569-A2.
XX
PD 19-MAR-2003.
XX
XX 21-MAR-2002; 2002EP-00006586.
PF
XX 14-SEP-2001; 2001JP-00328381.
PR 24-JAN-2002; 2002US-0350435P.
XX
XX (HELI-) HELIX RES INST.
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
XX
PI Isogai T, Sugiyama T, Otsuki T, Wakaratsu A, Sato H, Ishii S;
PI Yamamoto U, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX WPI; 2003-395539/38.
DR N-PSDB; ADA53741.
XX
XX New polynucleotides encoding full-length polypeptides, e.g. secretory
PT and/or membrane proteins, useful for developing medicines for diseases in
PT which the gene is involved, or as target molecules for gene therapy.
XX

XX Claim 14, SEQ ID NO 2948; 205bp; English.
XX
XX The present invention relates to novel human secretory or membrane
CC proteins (ADA54072-AD55710) and their coding sequences (ADA52433-
CC ADA54071). The coding sequences are useful in the gene therapy of
CC diseases caused by abnormalities of the proteins, e.g. cancer,
CC inflammatory diseases, osteoporosis or neurological disease.
XX
XX Sequence 176 AA;
SQ
Query Match 2.2%; Score 9; DB 6; Length 176;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 293 EDEDEDEED 301
Db 149 EDEDEDEED 157
RESULT 10
ABP28933
ID ABP28933 standard; protein; 202 AA.
AC ABP28933;
XX
XX 02-JUL-2002 (first entry)
DT
XX Streptococcus polypeptide SEQ ID NO 7042.
DE
XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial;
KM antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX
XX Streptococcus pyogenes.
OS
XX WO200234771-A2.
PN
XX 02-MAY-2002.
PD
XX 29-OCT-2001; 2001WO-GB004789.
XX
XX 27-OCT-2000; 2000GB-00026333.
PR 24-NOV-2000; 2000GB-00028727.
PR 07-MAR-2001; 2001GB-00005640.
XX
XX (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
XX Telford J, Masignani V, Margarit Y Rosi, Grandi G, Fraser C;
PI Tettein H;
XX
XX WPI; 2002-352536/38.
DR N-PSDB; ABN69564.
XX
XX New Streptococcus protein for the treatment or prevention of infection or
PT disease caused by Streptococcus bacteria, such as meningitis, and for
PT detecting a compound that binds to the protein.
XX
XX Claim 1; Page 3865; 4525pp; English.
PS
XX The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by

CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins
XX
XX Sequence 202 AA;
SQ
Query Match 2.2%; Score 9; DB 5; Length 202;
Best Local Similarity 100.0%; Pred. No. 5.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 293 EDEDEDEED 301
Db 189 EDEDEDEED 197
RESULT 11
ADD47645
ID ADD47645 standard; protein; 214 AA.
XX
XX ADD47645;
AC
XX 29-JAN-2004 (first entry)
DT
XX Human Protein P09429, SEQ ID NO 13341.
DE
XX Human; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; chronic constriction injury; CCI;
KW spared nerve injury; SNI; Chung.
XX
XX Homo sapiens.
OS
XX WO2003016475-A2.
PN
XX 27-FEB-2003.
PD
XX 14-AUG-2002; 2002WO-US025765.
XX
XX 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX
XX (GENO) GEN HOSPITAL CORP.
PA (FARB) BAYER AG.
XX
XX Woolf C, D'urso D, Befort K, Costigan M;
PI
XX WPI; 2003-268312/26.
DR GENBANK; P09429.
DR
XX
XX New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
PT
XX
XX Claim 1; Page; 1017pp; English.
PS
XX The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating

CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a human protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 214 AA;

Query Match 2.2%; Score 9; DB 7; Length 214;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 293 EDEDEDED 301
Db 187 EDEDEDED 195

RESULT 14
ADE57984
ID ADE57984 standard; protein; 214 AA.

XX ADE57984;

DT 29-JAN-2004 (first entry)

DE Human Protein P09429, SEQ ID NO 3852.

XX Human; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; chronic constriction injury; CCI;
KM spared nerve injury; SNI; Chung.

XX Homo sapiens.

PN WO2003016475-A2.

PD 27-FEB-2003.

PF 14-AUG-2002; 2002WO-US025765.

PR 14-AUG-2001; 2001US-0312147P.

PR 01-NOV-2001; 2001US-0346382P.

PR 26-NOV-2001; 2001US-0333347P.

XX (GEHO) GEN HOSPITAL CORP.

PA (FARB) BAYER AG.

PI Woolf C, D'urso D, Befort K, Costigan M;

DR WPI; 2003-268312/26.

XX GENBANK; P09429.

XX New composition comprising two or more isolated polypeptides, useful for

PT preparing a medicament for treating pain in an animal.

XX Claim 1; Page; 1017p; English.

XX The invention discloses a composition comprising two or more isolated rat

CC or human polynucleotides or a polynucleotide which represents a fragment,

CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a human protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 214 AA;

Query Match 2.2%; Score 9; DB 7; Length 214;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 293 EDEDEDED 301
Db 187 EDEDEDED 195

RESULT 15

ADE60732
ID ADE60732 standard; protein; 214 AA.

XX ADE60732;

DT 29-JAN-2004 (first entry)

DE Human Protein P09429, SEQ ID NO 6644.

XX Human; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; chronic constriction injury; CCI;
KM spared nerve injury; SNI; Chung.

XX Homo sapiens.

PN WO2003016475-A2.

PD 27-FEB-2003.

PF 14-AUG-2002; 2002WO-US025765.

PR 14-AUG-2001; 2001US-0312147P.

PR 01-NOV-2001; 2001US-0346382P.

PR 26-NOV-2001; 2001US-0333347P.

XX (GEHO) GEN HOSPITAL CORP.

PA (FARB) BAYER AG.

PI Woolf C, D'urso D, Befort K, Costigan M;

DR WPI; 2003-268312/26.

XX GENBANK; P09429.

XX New composition comprising two or more isolated polypeptides, useful for

PT preparing a medicament for treating pain in an animal.

XX Claim 1; Page: 1017pp; English.
PS
XX

CC The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (SNI), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a human protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX

SO Sequence 214 AA;

Query Match 2.2%; Score 9; DB 7; Length 214;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 293 EDEDEDEED 301
Db 187 EDEDEDEED 195

Search completed: July 22, 2004, 16:54:15
Job time : 56 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2004, 16:48:32 ; Search time 13 Seconds

(Without alignments)
1650.223 Million cell updates/sec

Title: US-09-831-804-3

Sequence: 1 MSSEDTKTSISSIISSSSSS.....PLVKARMDLLPNETSVISR 412

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 0

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|--------------|--------------------|
| 1 | 14 | 3.4 | 374 | TF3A_SCHPO | Q9ut15 schizosacch |
| 2 | 10 | 2.4 | 645 | 12235_MOUSE | Q61116 mus musculu |
| 3 | 9 | 2.2 | 191 | RPOB_STRPB | Q8nff7 streptococc |
| 4 | 9 | 2.2 | 191 | RPOB_STRPB | P58053 streptococc |
| 5 | 9 | 2.2 | 211 | HMTX_HUMAN | Q9ugv6 homo sapien |
| 6 | 9 | 2.2 | 214 | 1 HMG1_BOVIN | P10103 bos taurus |
| 7 | 9 | 2.2 | 214 | 1 HMG1_HUMAN | P09429 homo sapien |
| 8 | 9 | 2.2 | 333 | 1 KC2B_NEUCR | P12682 sus scrofa |
| 9 | 9 | 2.2 | 336 | 1 TRT_DROME | Q8eg12 neuropeptid |
| 10 | 9 | 2.2 | 571 | 1 YK18_YEAST | P19351 drosophila |
| 11 | 9 | 2.2 | 589 | 1 RGP1_MOUSE | P36076 saccharomyc |
| 12 | 9 | 2.2 | 606 | 1 CENB_CRIGR | P46061 mus musculu |
| 13 | 9 | 2.2 | 699 | 1 ECM2_HUMAN | P48988 cricetulus |
| 14 | 9 | 2.2 | 700 | 1 CH60_PLARG | Q94769 homo sapien |
| 15 | 9 | 2.2 | 758 | 1 YP58_YEAST | P34440 plasmodium |
| 16 | 9 | 2.2 | 830 | 1 YIP2_MOUSE | Q99299 saccharomyc |
| 17 | 9 | 2.2 | 104 | 1 HSP3_RAT | Q9e9e9 mus musculu |
| 18 | 8 | 1.9 | 115 | 1 NUFM_RAT | Q63622 rattus norv |
| 19 | 8 | 1.9 | 149 | 1 HMG1_VICFA | P40620 vicia faba |
| 20 | 8 | 1.9 | 180 | 1 HMG1_CRIGR | P07156 cricetulus |
| 21 | 8 | 1.9 | 214 | 1 HMG1_MOUSE | P07155 mus musculu |
| 22 | 8 | 1.9 | 344 | 1 YB1E_SCHPO | P87179 schizosacch |
| 23 | 8 | 1.9 | 359 | 1 IE68_EOGL | Q64953 eucalyptus |
| 24 | 8 | 1.9 | 377 | 1 IE68_PRIVK | P24627 pseudorabie |
| 25 | 8 | 1.9 | 402 | 1 PEKE_MOUSE | Q75381 homo sapien |
| 26 | 8 | 1.9 | 412 | 1 PEKE_MOUSE | Q55738 mus musculu |
| 27 | 8 | 1.9 | 417 | 1 FNA4_SPOFR | Q32686 spodoptera |
| 28 | 8 | 1.9 | 439 | 1 CRTG_HUMAN | P27797 homo sapien |
| 29 | 8 | 1.9 | 439 | 1 LIP1_DROME | Q46107 drosophila |
| 30 | 8 | 1.9 | 446 | 1 IRX2_HUMAN | Q9p657 homo sapien |
| 31 | 8 | 1.9 | 474 | 1 IRX2_MOUSE | P81066 mus musculu |
| 32 | 8 | 1.9 | 542 | 1 TUL1_HUMAN | Q00294 homo sapien |
| 33 | 8 | 1.9 | 542 | 1 TUL1_HUMAN | Q00294 homo sapien |

| | | | | | |
|----|---|-----|-----|---------------|---------------------|
| 34 | 8 | 1.9 | 547 | 1 IF37_MOUSE | Q70194 mus musculu |
| 35 | 8 | 1.9 | 584 | 1 YMB3_YEAST | Q04228 saccharomyc |
| 36 | 8 | 1.9 | 591 | 1 PAK4_HUMAN | Q96013 homo sapien |
| 37 | 8 | 1.9 | 628 | 1 DY11_MOUSE | Q88485 mus musculu |
| 38 | 8 | 1.9 | 628 | 1 HAP1_MOUSE | Q35668 mus musculu |
| 39 | 8 | 1.9 | 629 | 1 HAP1_RAT | P54256 rattus norv |
| 40 | 8 | 1.9 | 643 | 1 DY11_RAT | Q63100 rattus norv |
| 41 | 8 | 1.9 | 645 | 1 DY11_HUMAN | Q14576 homo sapien |
| 42 | 8 | 1.9 | 689 | 1 MEA_ARATH | Q65312 arabidopsis |
| 43 | 8 | 1.9 | 694 | 1 NUCL_CHICK | P15771 gallus galli |
| 44 | 8 | 1.9 | 695 | 1 APP2_MOUSE | Q06335 mus musculu |
| 45 | 8 | 1.9 | 701 | 1 UBF2_XENTIA | P25980 xenopus lae |

ALIGNMENTS

RESULT 1
TF3A_SCHPO STANDARD; PRT; 374 AA.
ID O9UT15;
AC 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
GN Transcription factor IIA (Factor A) (TFIIIA).
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A. AND FUNCTION.
RX MEDLINE=22082319; PubMed=12087160;
RA Schulman D.B., Seczer D.R.;
RT Identification and characterization of transcription factor IIA from
RL Schizosaccharomyces pombe.";
RL Nucleic Acids Res. 30:2772-2781(2002).
RN
RP
RC SOURCE FROM N.A.
RX STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajadream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Godle A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leacher S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Vancleave G., Aert R., Robben J., Grynolprez B.,
RA Welfens I., Vanstreels E., Rieger M., Schaefer M., Moller-ner S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehman H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeart F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucifant M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Renvelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cernutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shporkovskii G.V., Ussery D., Barrell B.G., Nurse P.;
RL Nature 415:871-880(2002).
-!- FUNCTION: Is required for correct transcription of 5S RNA genes by
transcription of the 5S ribosomal RNA gene.
-!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 CC EMBL: AY091590; AAM00046.1; -
 CC EMBL: AL132675; CAB59689.1; -
 CC PIR: T37676; T37676.
 CC HSSP: P07248; IARD.
 CC GeneDB: SPombe; SPAC14.09c; -
 CC InterPro: IPR007087; Znf_C2H2.
 CC Pfam: PF00096; zf-C2H2; 10.
 CC ProDom: PD000003; Znf_C2H2; 1.
 CC SMART: SM00355; Znf_C2H2; 9.
 CC PROSITE: PS00028; ZINC_FINGER_C2H2_1; 8.
 CC PROSITE: PS0157; ZINC_FINGER_C2H2_2; 9.
 CC Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
 KW RNA-binding; Repeat; Nuclear protein.
 FT ZN_FING 23 47 C2H2-TYPE.
 FT ZN_FING 53 77 C2H2-TYPE.
 FT ZN_FING 83 107 C2H2-TYPE.
 FT ZN_FING 113 138 C2H2-TYPE.
 FT ZN_FING 144 169 C2H2-TYPE.
 FT ZN_FING 204 226 C2H2-TYPE.
 FT ZN_FING 236 261 C2H2-TYPE.
 FT ZN_FING 267 291 C2H2-TYPE (ATYPICAL).
 FT ZN_FING 349 374 C2H2-TYPE.
 SQ SEQUENCE 374 AA; 43851 MM; 7469C701FF08FF6 CRC64;

Query Match 3.4%; Score 14; DB 1; Length 374;
 Best Local Similarity 100.0%; Pred. No. 1.4e-05;
 Matches 14; Conservativity 0; Mismatches 0; Indels 0; Gaps 0;

Oy 38 RPSLLEQHLRTHSN 51
 Db 36 RPSLLEQHLRTHSN 49

RESULT 2
 Z235_MOUSE STANDARD; PRT; 645 AA.
 AC Q61116; Q8C70; -
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Zinc finger protein 235 (Zinc finger protein 93) (Zfp-93).
 GN ZNF235 OR ZFP93.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CD-1; TISSUE=Testis;
 RX MEDLINE=96207307; PubMed=8617494;
 RA Shannon M., Ashworth L.K., Mucenski M.L., Lamerdin J.E., Branscomb E.,
 RA Stubbs L.;
 RT "Comparative analysis of a conserved zinc finger gene cluster on human
 RT chromosome 19q and mouse chromosome 7.";
 RL Genomics 33:112-120(1996).
 RU [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Skin, and Thymus;
 RX MEDLINE=22354683; PubMed=12466851;
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
 RA Nishida I., Osato N., Saito K., Suzuki H., Yamataka I., Kiyosawa H.,
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gotohori T.,
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
 RA Blake J.A., Bradt D., Brusic V., Chochia C., Corbani L.E., Cousins S.,

RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
 RA Grimmond S., Gustlinch S., Hirokawa N., Jackson I.J., Jarvis E.D.,
 RA Kanai A., Kawaji H., Kawasawa Y., Kedzielski R.M., King B.L.,
 RA Kurokawa A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
 RA Nagatake T., Numata K., Okido T., Pavan W.J., Pereira G., Pesole G.,
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S.,
 RA Ravasi T., Reed J.C., Reid D.J., Ring B.Z., Ringwald M.,
 RA Sandelin A., Schneider C., Sempke C.A., Setou M., Shimada K.,
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
 RA Wilming L.G., Wyszewski-Boris A., Yanagisawa M., Yang L.,
 RA Yuan Z., Zavalan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
 RA Hirokawa K., Kikawa T., Kono H., Nakamura M., Sakazume N., Sato K.,
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh N., Kagawa I.,
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
 RA Birney E., Hayashizaki Y.;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 RP [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
 RA Diatchenko L., Maruska K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mulhaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hultk S.W.,
 RA Vallat D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bonfield G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -1- FUNCTION: May function as a transcription factor.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
 CC FINGER PROTEINS.
 CC -1- SIMILARITY: Contains 1 KRAB domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 CC EMBL: U46186; AAB03529.1; -
 CC EMBL: AK028905; BAC26187.1; -
 CC EMBL: AK040325; BAC30565.1; -
 CC EMBL: BC003776; AAB03776.1; -
 CC HSSP: P25490; IZNM.
 CC MGD: MGI:107611; Zfp93.
 CC InterPro: IPR001909; KRAB.
 CC InterPro: IPR007087; Znf_C2H2.
 CC InterPro: IPR007086; Znf_C2H2_sub.
 CC Pfam: PF01352; KRAB. 1.
 CC Pfam: PF00096; zf-C2H2; 13.
 CC PRINTS: PR00048; ZINCFINGER.

DR ProDom: PD000003: Znf_C2H2; 11.
 DR SMART: SMO0349; KRAb_1.
 DR SMART: SMO0355; Znf_C2H2; 13.
 DR PROSITE; PS50805; KRAb_1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 13.
 DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 13.
 KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
 KW Nuclear protein; Repeat.
 FT DOMAIN 8
 FT ZN_FING 285 307 KRAb.
 FT ZN_FING 313 335 C2H2-TYPE.
 FT ZN_FING 341 363 C2H2-TYPE.
 FT ZN_FING 369 391 C2H2-TYPE.
 FT ZN_FING 397 419 C2H2-TYPE.
 FT ZN_FING 425 447 C2H2-TYPE.
 FT ZN_FING 453 475 C2H2-TYPE.
 FT ZN_FING 481 503 C2H2-TYPE.
 FT ZN_FING 509 531 C2H2-TYPE.
 FT ZN_FING 537 559 C2H2-TYPE.
 FT ZN_FING 565 587 C2H2-TYPE.
 FT ZN_FING 593 615 C2H2-TYPE.
 FT ZN_FING 621 643 C2H2-TYPE.
 FT CONFLICT 85 85 A -> P (IN REF. 2: BAC26187).
 SQ SEQUENCE 645 AA; 73012 MW; 7459CD140F5AF469 CRC64;

Query Match 2.4%; Score 10; DB 1; Length 645;
 Best Local Similarity 100.0%; Pred. No. 0.16;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 KPFGCVCGK 92
 DB 451 KPFGCVCGK 460

RESULT 3
 RPOE_STRP8
 ID RPOE_STRP8 STANDARD; PRT; 191 AA.
 AC QGNZF7;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable DNA-directed RNA polymerase delta subunit (RNAP delta factor).
 GN RPOE OR SPY18_1960.
 OS Streptococcus pyogenes (serotype M18).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=186103;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=MGAS8232 / Serotype M18;
 RX MEDLINE=21927593; PubMed=11917108;
 RA Smoot J.C., Barbican K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,
 RA Sylva G.L., Sturdevant D.B., Ricklefs S.M., Porcella S.F.,
 RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
 RA Kapur V., Daly J.A., Veasy D.G., Musser J.M.;
 RT "Genome sequence and comparative microarray analysis of serotype M18
 group A Streptococcus strains associated with acute rheumatic fever
 outbreaks";
 RT Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673(2002).
 RL [3]
 CC -!- FUNCTION: Participates in both the initiation and recycling phases
 of transcription. In the presence of the delta subunit, RNAP
 displays an increased specificity of transcription, a decreased
 affinity for nucleic acids, and an increased efficiency of RNA
 synthesis because of enhanced recycling (by similarity).
 CC -!- SUBUNIT: RNAP is composed of a core of 2 alpha, a beta and a beta'
 subunits. The core is associated with a delta subunit and one of
 several sigma factors (by similarity).
 CC -!- SIMILARITY: Belongs to the rpoB family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AE010100; AAL98451.1; -
 DR HAMAP; MF 00357; -; 1.
 DR InterPro; IPR007759; RNA pol delta.
 DR Pfam; PF05066; RNA pol_delta_1.
 KW DNA-directed RNA polymerase; Transcription; Complete proteome.
 FT DOMAIN 106 191 ASP/GLU-RICH (ACIDIC).
 SQ SEQUENCE 191 AA; 22250 MW; CD081EFA0BC7C329 CRC64;

Query Match 2.2%; Score 9; DB 1; Length 191;
 Best Local Similarity 100.0%; Pred. No. 0.49;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 293 EEDDEDEED 301
 DB 178 EEDDEDEED 186

RESULT 4
 RPOE_STRPY
 ID RPOE_STRPY STANDARD; PRT; 191 AA.
 AC P58053;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Probable DNA-directed RNA polymerase delta subunit (RNAP delta factor).
 GN RPOE OR SPY1895 OR SPY1895 OR SPY1895 OR SPS0233.
 OS Streptococcus pyogenes, and
 OS Streptococcus pyogenes (serotype M3).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1314, 198466;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=SP370 / ATCC 700294 / Serotype M1;
 RX MEDLINE=21192684; PubMed=11296296;
 RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
 RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
 RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
 RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
 RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes";
 RT Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
 RL [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=MGAS315 / Serotype M3;
 RX MEDLINE=22133808; PubMed=12122206;
 RA Beres S.B., Sylva G.L., Barbican K.D., Lei B., Hoff J.S.,
 RA Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,
 RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,
 RA Schlievert P.M., Musser J.M.;
 RT "Genome sequence of a serotype M3 strain of group A Streptococcus:
 RT phase-encoded toxins, the high-virulence phenotype, and clone
 emergence";
 RT Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).
 RL [3]
 CC -!- FUNCTION: Participates in both the initiation and recycling phases
 of transcription. In the presence of the delta subunit, RNAP
 displays an increased specificity of transcription, a decreased

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

```

CC affinity for nucleic acids, and an increased efficiency of RNA
CC synthesis because of enhanced recycling (by similarity).
CC -1- SUBUNIT: RNA is composed of a core of 2 alpha, a beta and a beta'
CC subunits. The core is associated with a delta subunit and one of
CC several sigma factors (by similarity).
CC -1- SIMILARITY: Belongs to the rpoC family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AE006614; AAK34603.1; -
CC EMBL; AE014167; AAM80240.1; ALT_INIT.
CC EMBL; AP005141; BAC63328.1; ALT_INIT.
CC HAMAP; MF_00357; -; 1.
CC InterPro; IPR007759; RNA_pol_delta.
CC Pfam; PF05066; RNA_pol_delta; 1.
CC DNA-directed RNA polymerase; Transcription; Complete proteome.
CC FT DOMAIN 118 191 ASP/GLU-RICH (ACIDIC).
CC SEQUENCE 191 AA; 22264 MW; D1781EFA0BC6926C CRC64;
-----
Query Match 2.2%; Score 9; DB 1; Length 191;
Best Local Similarity 100.0%; Pred. No. 0.49; Mismatches 0; Gaps 0;
Matches 9; Conservative 0; Indels 0;
-----
Qy 293 EDEDEDED 301
Db 178 EDEDEDED 186
-----
RESULT 5
HMXL HUMAN STANDARD; PRT; 211 AA.
ID HMXL_HUMAN
AC O9UGV6;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE High mobility group protein 1-like 10 (HMG-1L10).
OS Homo sapiens (Human).
GN Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RP SEQUENCE FROM N.A.
RX MEDLINE=20057165; PubMed=10591208;
RA Dunham I., Hunt A.R., Collins J.E., Brunikewich R., Beare D.M.,
RA Clump M., Smink L.U., Ainscough R., Almeida J.P., Babbage A.K.,
RA Baguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
RA Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,
RA Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
RA Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,
RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson B.,
RA Dharmi P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,
RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
RA Gilbert J.G.R., Goward M.E., Graham D.V., Griffiths M.N.D., Hall C.,
RA Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,
RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,
RA Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,
RA Matlyn I.D., Masinri-Mohammadi M., Matthews L.H., Mccann O.T.,
RA McElay J., McLaren S., McMurtry A.A., Milne S.A., Mortimore B.J.C.T.,
RA Ogden C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
RA Phillips S.H., Plumb R.W., Ramsey H., Ramsey Y., Rogers L., Rose M.T.,
RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
RA Soderlund C., Spragon L., Steward C.A., Suleston J.E., Swann R.M.,
RA Vandin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
RA Wright C.L., Hubbard K., Bentley D.R., Beck S., Rogers J., Shimizu N.,
RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,

```

```

RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,
RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,
RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,
RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
RA Zhan M., Zhang G., Chisoe S., Murray J., Miller N., Minx P.,
RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,
RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
RA Hinds K., Kemp K., Latreille P., Layman D., Ozerky P., Rohlfing T.,
RA Schaefer P., Walker C., Wamsley A., Wohlmann P., Fein K., Nelson J.,
RA Korf I., Bedell J.A., Hillier L.W., Mardis E., Waterston R.,
RA Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saito S.,
RA Badier M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,
RA Edelmann L., Kim U.J., Shibuya H., Simon M.I., Dumanski J.P.,
RA Peyraud M., Kedra D., Seroussi E., Pransson I., Tapia I., Bruder C.E.,
RA O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,
RA Khan A.S., Lane L., Tliahun Y., Wright H.,
RT "The DNA sequence of human chromosome 22."
RL Nature 402:489-495(1999).
CC -1- FUNCTION: Binds preferentially single-stranded DNA and unwinds
CC double stranded DNA (by similarity).
CC -1- SUBCELLULAR LOCATION: Nuclear (by similarity).
CC -1- SIMILARITY: Belongs to the HMG1/HMG2 protein family.
CC -1- SIMILARITY: Contains 2 HMG box domains.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; Z95115; CAB62951.1; -
CC HSSP; P07156; HNHN.
CC Genew; HGNC:494; HMG1L10.
CC InterPro; IPR000135; Highmobility_12.
CC InterPro; IPR000910; HMG_12_box.
CC Pfam; PF00505; HMG_box; 2.
CC PRINTS; PR00886; HIGHMOBILITY12.
CC SMART; SM00398; HMG; 2.
CC PROSITE; PS00353; HMG_BOX_1; 1.
CC PROSITE; PS00118; HMG_BOX_2; 2.
CC KW Nuclear protein; Chromosomal protein; DNA-binding; Repeat.
FT DNA_BIND 9 79 HMG_BOX 1.
FT DNA_BIND 95 163 HMG_BOX 2.
FT DOMAIN 186 211 ASP/GLU-RICH (ACIDIC).
FT SEQUENCE 211 AA; 24218 MW; 2A53BA2A6A6DF7CD CRC64;
-----
Query Match 2.2%; Score 9; DB 1; Length 211;
Best Local Similarity 100.0%; Pred. No. 0.53; Mismatches 0; Gaps 0;
Matches 9; Conservative 0; Indels 0;
-----
Qy 293 EDEDEDED 301
Db 188 EDEDEDED 196
-----
RESULT 6
HMG1 BOVIN STANDARD; PRT; 214 AA.
ID HMG1_BOVIN
AC P10103;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE High mobility group protein 1 (HMG-1).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OX NCBI_TaxID=9913;

```

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Holstein; TISSUE=Fetal thymus;
 RX MEDLINE=89057489; PubMed=3194213;
 RA Kaplan D.J., Duncan C.H.;
 RT "Full length cDNA sequence for bovine high mobility group 1 (HMG1)
 protein.";
 RL Nucleic Acids Res. 16:10375-10375(1988).
 RN [2]
 RP SEQUENCE OF 115-214 FROM N.A.
 RX MEDLINE=84128872; PubMed=6141822;
 RA Pentecost B., Dixon G.H.;
 RT "Isolation and partial sequence of bovine cDNA clones for the high-
 mobility-group protein (HMG-1).";
 RL Biosect. Rep. 4:49-57(1984).
 RN [3]
 RP SEQUENCE OF 1-36.
 RX MEDLINE=90306387; PubMed=2365081;
 RA Christen T., Bischoff M., Hobi R., Kuenzle C.C.;
 RT "High mobility group proteins 1 and 2 bind preferentially to
 brominated poly(dG-dC).poly(dG-dC) in the Z-DNA conformation but not
 to other types of Z-DNA.";
 RL FEBS Lett. 267:139-141(1990).
 CC -1- FUNCTION: Binds preferentially single-stranded DNA and unwinds
 CC double stranded DNA.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: Belongs to the HMG1/HMG2 protein family.
 CC -1- SIMILARITY: Contains 2 HMG box domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; X12796; CAA31284.1; -;
 DR EMBL; M26110; AAA30367.1; -;
 DR PIR; S01947; S01947.
 DR HSSP; P07156; INHN.
 DR InterPro; IPR000135; Highmobly_12.
 DR InterPro; IPR000910; HMG_12_box.
 DR Pfam; PF00505; HMG_box; 2.
 DR PRINTS; PR00886; HIGHMOBILITY12.
 DR SMART; SM00398; HMG; 2.
 DR PROSITE; PS00353; HMG_BOX_1; 1.
 DR PROSITE; PS0118; HMG_BOX_2; 2.
 KW Nuclear protein; Chromosomal protein; DNA-binding; Repeat.
 FT INIT MET 0
 FT DNA_BIND 8 78 HMG_BOX_1.
 FT DNA_BIND 94 162 HMG_BOX_2.
 FT DOMAIN 185 214 ASP/GLU-RICH (ACIDIC).
 FT CONFLICT 115 119 EHPGL->PGGAV (IN REF. 2).
 SQ SEQUENCE 214 AA: 24776 MW; B283A80FC7F0F433 CR664;
 Query Match 2.2%; Score 9; DB 1; Length 214;
 Best Local Similarity 100.0%; Pred. No. 0.54;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 293 EDEDEED 301
 DB 187 EDEDEED 195
 RESULT 7
 HMG1_HUMAN
 ID HMG1_HUMAN STANDARD; PRT; 214 AA.
 AC P09429;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE High mobility group protein 1 (HMG-1).

GN HMG1 OR HMG1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 ON NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89160247; PubMed=292262;
 RA Men L., Huang J.K., Johnson B.H., Reek G.R.;
 RT "A human placental cDNA clone that encodes nonhistone chromosomal
 protein HMG-1.";
 RL Nucleic Acids Res. 17:1197-1214(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96299787; PubMed=8661151;
 RA Ferrari S., Finelli P., Rocchi M., Bianchi M.E.;
 RT "The active gene that encodes human high mobility group 1 protein
 (HMG1) contains introns and maps to chromosome 13.";
 RL Genomics 35:367-371(1996).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Cervix, and Testis;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loggiano N.A., Peters R.J., Abramson R.D., Mollary S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Bullyk S.W.,
 RA Villalon D.K., Murray D.K., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalms D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Maira W.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [4]
 RP SEQUENCE OF 57-64 AND 112-126.
 RC TISSUE=Breast carcinoma;
 RX MEDLINE=97295304; PubMed=9150946;
 RA Rasmussen R.K., Ji H., Eddes J.S., Moritz R.L., Reid G.E.,
 RA Simpson R.J., Dorow D.S.;
 RT "Two-dimensional electrophoretic analysis of human breast carcinoma
 RT proteins: mapping of proteins that bind to the SH3 domain of mixed
 RT lineage kinase MLK2.";
 RL Electrophoresis 18:588-598(1997).
 CC -1- FUNCTION: Binds preferentially single-stranded DNA and unwinds
 CC double stranded DNA.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: Belongs to the HMG1/HMG2 protein family.
 CC -1- SIMILARITY: Contains 2 HMG box domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; X12597; CAA31110.1; -;
 DR EMBL; U51677; AAB08987.1; -;
 DR EMBL; BC003378; AAB03378.1; -;
 DR EMBL; BC030981; AAB030981.1; -;
 DR PIR; S02826; S02826.
 DR HSSP; P07156; INHN.

```

DR Genew; HGNC:4983; HMG1.
DR MIM; 163905; -.
DR GO; GO:0000228; C:nuclear chromosome; NAS.
DR GO; GO:0003697; P:single-stranded DNA binding; NAS.
DR GO; GO:0006268; P:DNA unwinding; NAS.
DR InterPro; IPR000135; Hichmobility_12.
DR InterPro; IPR000910; HMG_12_box.
DR Pfam; PF00505; HMG_box; 2.
DR PRINTS; PR00886; HIGHMOBILITY12.
DR SMART; SM00398; HMG; 2.
DR PROSITE; PS00353; HMG_BOX_1; 1.
DR PROSITE; PS50118; HMG_BOX_2; 2.
DR Nuclear protein; Chromosomal protein; DNA-binding; Repeat.
KW INIT_MET 0
FT DNA_BIND 8 78 HMG_BOX_1.
FT DNA_BIND 94 162 HMG_BOX_2.
FT DOMAIN 185 214 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 214 AA; 24762 MW; B3C6A80FC7F0FA33 CRC64;

Query Match
Best Local Similarity 100.0%; Score 9; DB 1; Length 214;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 293 EDEDEDED 301
DB 187 EDEDEDED 195

RESULT 8
HMG1_PIG STANDARD; PRT; 214 AA.
ID ID
AC P12682;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE High mobility group protein 1 (HMG-1).
GN HMG1 OR HMG1.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89050965; PubMed=119113;
RA Tsuda K.-I., Kikuchi M., Mori K., Waga S., Yoshida M.;
RT "Primary structure of non-histone protein HMG1 revealed by the
RT nucleotide sequence."
RL Biochemistry 27:6159-6163(1988).
CC -1- FUNCTION: Bands preferentially single-stranded DNA and unwinds
CC -1- double stranded DNA.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: Belongs to the HMG1/HMG2 protein family.
CC -1- SIMILARITY: Contains 2 HMG box domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M21683; AAA31050.1; -.
CC PIR; A28897; A28897.
CC HSSP; P07155; 1AAB.
CC InterPro; IPR000135; Hichmobility_12.
CC InterPro; IPR000910; HMG_12_box.
CC Pfam; PF00505; HMG_box; 2.
CC PRINTS; PR00886; HIGHMOBILITY12.
CC SMART; SM00398; HMG; 2.
CC PROSITE; PS00353; HMG_BOX_1; 1.
CC PROSITE; PS50118; HMG_BOX_2; 2.
KW Nuclear protein; Chromosomal protein; DNA-binding; Repeat.

```

```

FT INIT_MET 0 0
FT DNA_BIND 8 78 HMG_BOX_1.
FT DNA_BIND 94 162 HMG_BOX_2.
FT DOMAIN 185 214 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 214 AA; 24785 MW; B29C8A32D8D2C933 CRC64;

Query Match
Best Local Similarity 100.0%; Score 9; DB 1; Length 214;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 293 EDEDEDED 301
DB 187 EDEDEDED 195

RESULT 9
KC2B_NEUCR STANDARD; PRT; 333 AA.
ID ID
AC O8TG12;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Casein kinase II beta chain 1 (CK II beta 1).
GN CKB1.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21956574; PubMed=11959847;
RA Yang Y., Cheng P., Liu Y.;
RT "Regulation of the Neurospora circadian clock by casein kinase II."
RL Genes Dev. 16:994-1006(2002).
CC -1- FUNCTION: Plays a complex role in regulating the basal catalytic
CC activity of the alpha subunit (By similarity).
CC -1- SUBUNIT: Heterotrimer of 2 alpha subunits, a beta 1 subunit and
CC a beta 2 subunit.
CC -1- PTM: Phosphorylated by alpha chain (By similarity).
CC -1- SIMILARITY: Belongs to the casein kinase 2 beta chain family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF494377; AA014625.1; -.
CC InterPro; IPR00704; CAS_kinase_II.
CC Pfam; PF01214; CK II beta; 1.
CC PRINTS; PR00472; CASNKINASEII.
CC ProDom; PD003829; CAS_kinase_II; 1.
CC PROSITE; PS01101; CK2_BETA; 1.
KW Transferase; Serine/threonine-protein kinase; Phosphorylation.
SQ SEQUENCE 333 AA; 37107 MW; 1744C2D1F7E1D67C CRC64;

Query Match
Best Local Similarity 100.0%; Score 9; DB 1; Length 333;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 293 EDEDEDED 301
DB 68 EDEDEDED 76

RESULT 10
TBT_DROME STANDARD; PRT; 396 AA.
ID ID
AC P19351;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)

```

DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Troponin T, skeletal muscle (upheld protein) (Intended thorax
 DE protein).
 GN UP OR INT.
 OS Drosophila melanogaster (Fruit fly).
 CC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Asynchronous muscle;
 RX MEDLINE=89141761; PubMed=2852258;
 RA Bullard B., Leonard K., Larkins A., Butcher G., Karlik C.,
 RA Fryberg E.A.;
 RT "Troponin of asynchronous flight muscle.";
 RL J. Mol. Biol. 204:621-637(1988).
 RP [2]
 RN REVISIONS, SEQUENCE FROM N.A.
 RC STRAIN=Canton-S;
 RX MEDLINE=91080155; PubMed=2124273;
 RA Fryberg E.A., Fryberg C.C., Beall C., Saville D.L.;
 RT "Drosophila melanogaster troponin-T mutations engender three distinct
 RT syndromes of myofibrillar abnormalities.";
 RL J. Mol. Biol. 216:657-675(1990).
 CC -!- FUNCTION: Troponin T is the tropomyosin-binding subunit of
 CC troponin, the thin filament regulatory complex which confers
 CC calcium-sensitivity to striated muscle actomyosin ATPase activity.
 CC Mutations in troponin T engender three distinct syndromes of
 CC myofibrillar abnormalities.
 CC -!- SIMILARITY: Belongs to the troponin T family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X54504; CAA38366.1; -;
 DR PIR; S13251; S13251.
 DR Flybase; FBgn0004169; up.
 DR InterPro; IPR001978; Troponin.
 DR Pfam; PF00992; Troponin; 1.
 KW Muscle protein.
 KW DOMAIN
 FT 341 396 ASP/GLU-RICH (HIGHLY ACIDIC).
 SQ SEQUENCE 396 AA; 47333 MW; 3CAAA8AA8943C23 CRC64;
 Query Match 2.2%; Score 9; DB 1; Length 396;
 Best Local Similarity 100.0%; Pred. No. 0.93;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 293 EDEDEDED 301
 DB 358 EDEDEDED 366
 RESULT 11
 YK18 YEAST
 ID YK18 YEAST STANDARD; PRT; 571 AA.
 AC P36076;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical 65.2 kDa protein in MIF2-CYR2 intergenic region.
 GN YK1088W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4912;
 RN [1]
 RP SEQUENCE FROM N.A.

RA Pohl T.M., Pohl F.M.;
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: TO YEAST AND C.TROPICALIS S182/HAL3. SOME, TO
 CC A.THALIANA HAL3A AND HAL3B.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z28088; CAA81926.1; -;
 DR PIR; S37913; S37913.
 DR HSP; Q9SWE5; IE20.
 DR GerMolLine; I39844; -;
 DR SGD; S0001571; YK1088W.
 DR InterPro; IPR003382; Flavoprotein.
 DR Pfam; PF02441; Flavoprotein; 1.
 KW Hypothetical protein.
 FT DOMAIN 508 570
 SQ SEQUENCE 571 AA; 65238 MW; 9C674C2394EFCB8 CRC64;
 Query Match 2.2%; Score 9; DB 1; Length 571;
 Best Local Similarity 100.0%; Pred. No. 1.3;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 293 EDEDEDED 301
 DB 558 EDEDEDED 566
 RESULT 12
 RGP1 MOUSE
 ID RGP1 MOUSE STANDARD; PRT; 589 AA.
 AC P46061; Q60801;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Ran GTPase-activating protein 1.
 GN RANGAP1 OR FUG1.
 GN Mus musculus (Mouse).
 OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94148219; PubMed=8314081;
 RA Degregori J., Ruse A., von Melchner H., Rayburn H.,
 RA Priyaranjan P., Jenkins N.A., Copeland N.G., Rulley H.E.;
 RT "A murine homolog of the yeast RNL1 gene is required for
 RT postimplantation development.";
 RL Genes Dev. 8:265-276(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BA1B/c;
 RX MEDLINE=95198731; PubMed=7891706;
 RA Ren W., Villamarin A., Shih A., Coutavas E., Moore M.S.,
 RA Locurcio M., Clarke V., Oppenheim J.D., D'Eustachio P., Rush M.G.;
 RT "Separate domains of the Ran GTPase interact with different factors
 RT to regulate nuclear protein import and RNA processing.";
 RL Mol. Cell. Biol. 15:2117-2124(1995).
 CC -!- FUNCTION: GTPase activator for the nuclear Ras-related regulatory
 CC protein Ran, converting it to the putatively inactive GDP-bound
 CC state. Required for postimplantation development.
 CC SUBUNIT: Homodimer. Forms a tight complex in association with
 CC RANBP2 and the ubiquitin-conjugating enzyme E2 (UBC9) (By
 CC similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- PTM: Sumoylated.
 CC -!- SIMILARITY: Contains 6 leucine-rich (LRR) repeats.
 CC -!- SIMILARITY: To fungal RNL1.

```

CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
```

EMBL; U08110; AAA17681.1; -

EMBL; U0857; AAB60517.1; -

PIR; A36983; A36983.

PIR; T52070; T52070.

PDB; 1KE8; 13-FEB-02.

MD; MG1:103071; Rangapl.

InterPro; IPR001611; LRR.

InterPro; IPR007091; LRR_FinH.

Pfam; PF00560; LRR; 1.

GPase activation; Repeat; Leucine-rich repeat; 3D-structure.

REPEAT 48 71 LRR 1.

REPEAT 111 134 LRR 2.

REPEAT 207 230 LRR 3.

REPEAT 235 258 LRR 4.

REPEAT 292 315 LRR 5.

REPEAT 320 343 LRR 6.

DOMAIN 359 399 ASP/GDU-RICH (HIGHLY ACIDIC).

CONFLICT 181 LRR 1.

CONFLICT 413 413 S -> L (IN REF. 2).

SEQUENCE 589 AA; 63615 MW; 60FAA68B60B9669C CRC64;

Query Match 2.2%; Score 9; DB 1; Length 589;
Best Local Similarity 100.0%; Pred. No. 1.3; Mismatches 0; Gaps 0;
Matches 9; Conservative 0; Indels 0;

QY 293 EDEDEDED 301
Db 378 EDEDEDED 386

RESULT 13
CENB_CRIGR STANDARD; PRT; 606 AA.
ID_CENB_CRIGR STANDARD; PRT; 606 AA.
AC P48988;
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Major centromere autoantigen B (Centromere protein B) (CENP-B).
GN CENPB.
OS Cricetus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Criceulius.
OX NCBI_TaxID=10029;
OX [1]
RN SEQUENCE FROM N.A.
RP TISSUE=Ovary;
RX MEDLINE=96234058; PubMed=8652663;
RA Bejarano L.A., Valdivia M.M.;
RT "Molecular cloning of a nontransfected gene for the hamster centromere
RT antigen CENP-B";
RL Biochim. Biophys. Acta 1307:21-25(1996).
CC -i- FUNCTION: Interacts with centromeric heterochromatin in
CC chromosomes and binds to a specific subset of aliphoid satellite
CC DNA, called the CENP-B box. May organize arrays of centromere
CC satellite DNA into a higher order structure which then directs
CC centromere formation and kinetochore assembly in mammalian
CC chromosomes (By similarity).
CC -i- SUBUNIT: Homodimer (By similarity).
CC -i- SUBCELLULAR LOCATION: Nuclear.
CC -i- SIMILARITY: Contains 1 CENB domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

```

CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
```

EMBL; U0951; AAB06494.1; -

PIR; S70358; S70358.

HSSP; P07199; 1BM6.

InterPro; IPR004875; CENP-B.

InterPro; IPR006895; CENP-B_N.

InterPro; IPR006600; CENPB.

Pfam; PF04218; CENP-B_N; 1.

Pfam; PF03184; DDR; 1.

SMART; SM00674; CENPB; 1.

Chromosomal protein; Nuclear protein; DNA-binding; Centromere.

DNA BIND 1 125 BY SIMILARITY

DOMAIN 404 471 GDU-RICH (ACIDIC).

DOMAIN 510 545 ASP/GDU-RICH (ACIDIC).

SEQUENCE 606 AA; 66407 MW; 63DDEEB551D48E CRC64;

Query Match 2.2%; Score 9; DB 1; Length 606;
Best Local Similarity 100.0%; Pred. No. 1.4; Mismatches 0; Gaps 0;
Matches 9; Conservative 0; Indels 0;

QY 293 EDEDEDED 301
Db 528 EDEDEDED 536

RESULT 14
ECM2 HUMAN STANDARD; PRT; 699 AA.
ID_ECM2 HUMAN STANDARD; PRT; 699 AA.
AC O94769;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Extracellular matrix protein 2 precursor (Matrix glycoprotein
DE SCI/ECM2).
GN ECM2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
OX [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=9909324; PubMed=9790758;
RA Nishitani U., Tanaka T., Nakamura Y.;
RT "Identification of a novel gene (ECM2) encoding a putative
RT extracellular matrix protein expressed predominantly in adipose and
RT female-specific tissues and its chromosomal localization to 9q22.3";
RL Genomics 52:378-381(1998).
CC -i- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -i- TISSUE SPECIFICITY: Expressed predominantly in adipose tissue as
CC well as female-specific organs such as mammary gland, ovary, and
CC uterus.
CC -i- SIMILARITY: Belongs to the small leucine-rich proteoglycan (SLRP)
CC family. Class I subfamily.
CC -i- SIMILARITY: Contains 1 WFRC domain.
CC -i- SIMILARITY: Contains 12 leucine-rich (LRR) repeats.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

DR GO: GO:0005178; F: integrin binding; TAS.
 DR GO: GO:0007160; P: cell-matrix adhesion; TAS.
 DR InterPro: IPR001611; IRR.
 DR InterPro: IPR003591; IRR_LYP.
 DR InterPro: IPR001007; VWF_C.
 DR Pfam: PF00560; IRR; 10.
 DR Pfam: PF00093; VWC; 1.
 DR PRINTS: PR00019; LEORICHRPT.
 DR SMART: SM00214; VWC; 1.
 DR PROSITE: PS01208; VWF_C_1; 1.
 DR PROSITE: PS0184; VWF_C_2; 1.
 DR Leucine-rich repeat; Repeat; Glycoprotein; Extracellular matrix;
 KW Signal.
 FT SIGNAL 1 20
 FT CHAIN 21 699
 FT DOMAIN 101 158
 FT REPEAT 335 355
 FT REPEAT 356 379
 FT REPEAT 382 405
 FT REPEAT 406 426
 FT REPEAT 427 450
 FT REPEAT 451 476
 FT REPEAT 477 497
 FT REPEAT 524 547
 FT REPEAT 596 620
 FT REPEAT 621 643
 FT REPEAT 644 672
 FT REPEAT 673 699
 FT DOMAIN 270 281
 FT SITE 294 296
 FT CARBOHYD 378 378
 FT CARBOHYD 449 449
 FT CARBOHYD 506 506
 SQ SEQUENCE 699 AA; 79789 MW; E4E76A40A5C2742 CRC64;

Query Match 2.2%; Score 9; DB 1; Length 699;
 Best Local Similarity 100.0%; Pred. No. 1.5;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 293 EDEDEDEED 301
 Db 283 EDEDEDEED 291

RESULT 15
 CH60 PLAFG STANDARD; PRT; 700 AA.
 AC P34910;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Chaperonin CPN60, mitochondrial precursor.
 OS Plasmodium falciparum (isolate FCR-3 / Gambia).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5838;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94359534; PubMed=7915821;
 RA Hollaway S.P., Min W., Inselburg J.I.;
 RT "Isolation and characterization of a chaperonin-60 gene of the human
 malaria parasite Plasmodium falciparum.";
 RL Mol. Biochem. Parasitol. 64:25-32(1994).
 CC -1- FUNCTION: Implicated in mitochondrial protein import and
 macro-molecular assembly. May facilitate the correct folding of
 imported proteins. May also prevent misfolding and promote the
 refolding and proper assembly of unfolded polypeptides generated
 under stress conditions in the mitochondrial matrix.
 CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix (By similarity).
 CC -1- SIMILARITY: Belongs to the chaperonin (HSP60) family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X75420; CA53172.1; -.
 DR PIR: S38426; S38426.
 DR HSP: P06139; 1GRL.
 DR InterPro: IPR001844; Chaperonin Cpn60.
 DR InterPro: IPR002423; Cpn60/TCF-1.
 DR InterPro: IPR008950; GroEL-ATPase.
 DR Pfam: PF00118; cpn60_TCF1; 1.
 DR PRINTS: PR00298; CHAPERONIN60.
 DR PROSITE: PS00304; TCOMPLEXTCP1.
 DR Chaperone; ATP-binding; Transit peptide; Mitochondrion; Heat shock.
 KW TRANSIT 1 9
 FT CHAIN 10 700
 FT DOMAIN 643 700
 SQ SEQUENCE 700 AA; 79444 MW; AE1DB362B66D382 CRC64;

Query Match 2.2%; Score 9; DB 1; Length 700;
 Best Local Similarity 100.0%; Pred. No. 1.6;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 293 EDEDEDEED 301
 Db 674 EDEDEDEED 682

Search completed: July 22, 2004, 16:54:39
 Job time : 15 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2004, 16:51:37 ; Search time 45 Seconds

(without alignments)
2888.744 Million cell updates/sec

Title: US-09-831-804-3

Perfect score: 412

Sequence: 1 MESSDETKSISSLISSSSSS.....PLVKARMDLPNETSVISR 412

Scoring table: OLIGO
Gapco 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|--------|--------------------|
| 1 | 10 | 2.4 | 546 | 11 | Q61491 | Q61491 mus musculu |
| 2 | 10 | 2.4 | 546 | 11 | Q8VENO | Q8VENO mus musculu |
| 3 | 10 | 2.4 | 798 | 5 | O45181 | O45181 caenorhabdi |
| 4 | 10 | 2.4 | 6473 | 5 | Q8IKH9 | Q8IKH9 plasmodium |
| 5 | 9 | 2.2 | 163 | 11 | Q8BIC5 | Q8BIC5 mus musculu |
| 6 | 9 | 2.2 | 208 | 11 | Q8QVZ1 | Q8QVZ1 mus musculu |
| 7 | 9 | 2.2 | 211 | 4 | Q9NQJ4 | Q9NQJ4 homo sapien |
| 8 | 9 | 2.2 | 215 | 11 | Q88612 | Q88612 spalax leuc |
| 9 | 9 | 2.2 | 215 | 11 | Q9QWY6 | Q9QWY6 spalax leuc |
| 10 | 9 | 2.2 | 215 | 11 | Q9QX40 | Q9QX40 mus musculu |
| 11 | 9 | 2.2 | 354 | 11 | Q8BIJ2 | Q8BIJ2 mus musculu |
| 12 | 9 | 2.2 | 372 | 10 | Q9SZ12 | Q9SZ12 arabidopsis |
| 13 | 9 | 2.2 | 382 | 10 | Q38809 | Q38809 arabidopsis |
| 14 | 9 | 2.2 | 390 | 13 | Q9DFB9 | Q9DFB9 ictalurus p |
| 15 | 9 | 2.2 | 417 | 4 | Q86YR3 | Q86YR3 homo sapien |
| 16 | 9 | 2.2 | 435 | 3 | Q06523 | Q06523 saccharomyc |

| | | | | | | |
|----|---|-----|------|----|--------|---------------------|
| 17 | 9 | 2.2 | 518 | 11 | Q80W88 | Q80W88 mus musculu |
| 18 | 9 | 2.2 | 575 | 5 | Q81205 | Q81205 plasmodium |
| 19 | 9 | 2.2 | 589 | 11 | Q91YS2 | Q91YS2 mus musculu |
| 20 | 9 | 2.2 | 589 | 11 | Q8C2E3 | Q8C2E3 mus musculu |
| 21 | 9 | 2.2 | 589 | 11 | Q7TWM1 | Q7TWM1 mus musculu |
| 22 | 9 | 2.2 | 639 | 4 | Q81ZG1 | Q81ZG1 homo sapien |
| 23 | 9 | 2.2 | 639 | 4 | Q81UR6 | Q81UR6 homo sapien |
| 24 | 9 | 2.2 | 693 | 4 | Q7Z3D0 | Q7Z3D0 homo sapien |
| 25 | 9 | 2.2 | 702 | 5 | Q810V3 | Q810V3 plasmodium |
| 26 | 9 | 2.2 | 707 | 11 | Q99XK0 | Q99XK0 mus musculu |
| 27 | 9 | 2.2 | 774 | 11 | Q8R1Q5 | Q8R1Q5 mus musculu |
| 28 | 9 | 2.2 | 774 | 11 | Q8B7I8 | Q8B7I8 mus musculu |
| 29 | 9 | 2.2 | 775 | 11 | Q9D6C5 | Q9D6C5 mus musculu |
| 30 | 9 | 2.2 | 775 | 11 | Q8BFO7 | Q8BFO7 mus musculu |
| 31 | 9 | 2.2 | 792 | 12 | Q9YTL7 | Q9YTL7 ateline her |
| 32 | 9 | 2.2 | 1012 | 16 | Q7ULN3 | Q7ULN3 rhodoptrell |
| 33 | 9 | 2.2 | 1127 | 5 | Q81S17 | Q81S17 dictyostel |
| 34 | 9 | 2.2 | 1312 | 3 | Q8WZV2 | Q8WZV2 neurospora |
| 35 | 9 | 2.2 | 1404 | 13 | Q08757 | Q08757 gallus gall |
| 36 | 9 | 2.2 | 2075 | 5 | Q81HR5 | Q81HR5 plasmodium |
| 37 | 8 | 1.9 | 76 | 2 | P72352 | P72352 staphylococ |
| 38 | 8 | 1.9 | 77 | 16 | Q8YP97 | Q8YP97 arabidena sp |
| 39 | 8 | 1.9 | 106 | 11 | Q9D822 | Q9D822 mus musculu |
| 40 | 8 | 1.9 | 116 | 11 | Q9DAB6 | Q9DAB6 mus musculu |
| 41 | 8 | 1.9 | 119 | 10 | Q94190 | Q94190 atropa bell |
| 42 | 8 | 1.9 | 159 | 17 | Q9YD24 | Q9YD24 aeropyrum p |
| 43 | 8 | 1.9 | 176 | 2 | Q9EZB9 | Q9EZB9 staphylococ |
| 44 | 8 | 1.9 | 176 | 2 | Q848T9 | Q848T9 enterococc |
| 45 | 8 | 1.9 | 180 | 2 | Q9A103 | Q9A103 enterococc |

ALIGNMENTS

RESULT 1
Q61491 PRELIMINARY; PRT; 546 AA.
ID Q61491
AC Q61491;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE DNA-binding protein.
GN ZEP97.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95137393; PubMed=7835705;
RA Wick M.J., Ann D.K., Lee N.M., Ioh H.H.;
RT "Isolation of a cDNA encoding a novel zinc-finger protein from
neuroblastoma x glioma NG108-15 cells.";
RL Gene 152:227-232(1995).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL; L20450; AAA65196.1; -;
DR PIR; I49636; I49636.
DR HSSP; P08047; 1SP2.
DR MGD; MGI:105921; Zfp97.
DR GO; GO:0005634; C:nucleus; IDA.
DR InterPro; IPR001909; KRAb.
DR InterPro; IPR007087; Znf_C2H2.
DR InterPro; IPR007086; Znf_C2H2_sub.
DR Pfam; PF01352; KRAb; 1.
DR Pfam; PF00096; Zf-C2H2; 13.
DR PRINTS; PR00048; ZINCFINGER.
DR ProDom; PD000003; Znf_C2H2; 11.
DR SMART; SM00349; KRAb; 1.
DR SMART; SM00355; Znf_C2H2; 13.
DR PROSITE; PS50805; KRAb; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 13.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 14.
KW DNA-binding; Metal-binding; Nuclear protein; Zinc; Zinc-finger.

SQ SEQUENCE 546 AA; 64148 MW; BEA735BAF5C5528B CRC64;
 Query Match 2.4%; Score 10; DB 11; Length 546;
 Best Local Similarity 100.0%; Pred. No. 0.58;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 LKRHEITHTK 109
 DB 505 LKRHEITHTK 514

RESULT 2
 ID O8VENO PRELIMINARY; PRT; 546 AA.
 AC O8VENO;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, last annotation update)
 DE Similar to zinc finger protein 97.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC018101; AAH18101.1; -.
 DR GO; GO:0005622; C:intracellular; IEA.
 DR GO; GO:0003676; F:nucleic acid binding; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR001909; KRAB
 DR InterPro; IPR007087; ZNF_C2H2.
 DR InterPro; IPR007086; ZNF_C2H2_sub.
 DR Pfam; PF01352; KRAB; 1.
 DR Pfam; PF00096; zf_C2H2; 13.
 DR PRINTS; PR00048; ZINC_FINGER.
 DR ProDom; PD000003; ZNF_C2H2; 10.
 DR SMART; SM00349; KRAB; 1.
 DR SMART; SM00355; ZNF_C2H2; 14.
 DR PROSITE; PSS0805; KRAB; 1.
 DR PROSITE; PSS00028; ZINC_FINGER_C2H2_1; 13.
 DR PROSITE; PSS0157; ZINC_FINGER_C2H2_2; 14.
 KM Metal-binding; Zinc; Zinc-finger
 KM SEQUENCE 546 AA; 63973 MW; 092E26D554609D5D CRC64;
 SQ

Query Match 2.4%; Score 10; DB 11; Length 546;
 Best Local Similarity 100.0%; Pred. No. 0.58;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 LKRHEITHTK 109
 DB 505 LKRHEITHTK 514

RESULT 3
 ID 045181 PRELIMINARY; PRT; 798 AA.
 AC 045181;
 DT 01-JUN-1998 (TREMBlrel. 06, Created)
 DT 01-JUN-1998 (TREMBlrel. 06, last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, last annotation update)
 DE Hypothetical protein.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nemata; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA None;

RT "Genome sequence of the nematode C. elegans: a platform for
 RT investigating biology. The C. elegans Sequencing Consortium.";
 RT Science 282:2012-2018(1998).
 RN [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=Bristol N2;
 RC Pullon B., Hawkins J., Gattung S., Wohlmann P., Elliott G.;
 RT "The sequence of C. elegans cosmid K07H8.";
 RT Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RN SEQUENCE FROM N.A.
 RP STRAIN=Bristol N2;
 RA Waterston R.;
 RT "Direct Submission.";
 RT Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF047659; AAC04430.1; -.
 DR PIR; T33022; T33022.
 DR WormPep; K07H8.10; CE18030.
 DR GO; GO:0003676; F:nucleic acid binding; IEA.
 DR InterPro; IPR005054; RNA_rec_mot.
 DR Pfam; PF00076; rrm; 2.
 DR SMART; SM00360; RRM; 2.
 DR PROSITE; PSS0102; RRM; 1.
 KM Hypothetical protein.
 SQ SEQUENCE 798 AA; 86819 MW; 6209063CD13FAB26 CRC64;
 SQ

Query Match 2.4%; Score 10; DB 5; Length 798;
 Best Local Similarity 100.0%; Pred. No. 0.82;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 292 VEEDEDEED 301
 DB 287 VEEDEDEED 296

RESULT 4
 ID O8IKH9 PRELIMINARY; PRT; 6473 AA.
 AC O8IKH9;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, last sequence update)
 DT 01-JUN-2003 (TREMBlrel. 24, last annotation update)
 DE Dynein beta chain, putative.
 GN PF14_0626.
 OS Plasmodium falciparum (isolate 3D7).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OC NCBI_TaxID=36329;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=3D7;
 RX MEDLINE=2255705; PubMed=12368864;
 RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
 RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
 RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
 RA Chan M.-S., Nene V., Shallow S.D., Suh B., Peterson J., Anguioni S.,
 RA Pereira M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
 RA Martin D.M.A., Fairlamb A.H., Fraumholz M.J., Roos D.S., Ralph S.A.,
 RA McPadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
 RA Venner J.C., Carnuci B.J., Hoffman S.L., Newbold C., Davis R.W.,
 RA Fraser C.M., Barrell B.;
 RT "Genome sequence of the human malaria parasite Plasmodium
 RT falciparum.";
 RL Nature 419:498-511(2002).
 DR EMBL; AF014826; AAN37239.1; -.
 DR GO; GO:0030286; C:dyein complex; IEA.
 DR GO; GO:0008567; F:dyein ATPase activity; IEA.
 DR GO; GO:0003777; F:microtubule motor activity; IEA.
 DR GO; GO:0007018; F:microtubule-based movement; IEA.
 DR InterPro; IPR004273; Dynein heavy.
 DR Pfam; PF03028; Dynein heavy; 1.
 SQ SEQUENCE 6473 AA; 771748 MW; BFC848CD45D5D2F8 CRC64;
 SQ

Query Match 2.4%; Score 10; DB 5; Length 6473;

Best Local Similarity 100.0%; Pred. No. 5.2;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 292 VEDEDEDED 301
Db 5443 VEDEDEDED 5452

RESULT 5

Q8B1C5 PRELIMINARY; PRT; 163 AA.
AC Q8B1C5;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-OCT-2003 (TREMBLrel. 25, last sequence update)
DE Weakly similar to BA393J16.3.
GN BA114266.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN [1]
RE SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium.
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK089532; BAC40916.1; -
DR MGD; MGI:2143362; BB114266.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; Zf-C2H2; 5.
DR SMART; SM00355; Znf_C2H2; 5.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_2; 5.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 5.
SQ SEQUENCE 163 AA; 19461 MW; ALAFDHC074C04 CRC64;

Query Match 2.2%; Score 9; DB 11; Length 163;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 64 KAFPRKSHL 72
Db 29 KAFPRKSHL 37

RESULT 6

O80YZ1 PRELIMINARY; PRT; 208 AA.
AC O80YZ1;
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, last annotation update)
DE BML68F16.1 (Novel protein similar to high-mobility group box 1
DE (Hmgbl1)).
GN BML68F16.1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN [1]
RE SEQUENCE FROM N.A.
RC North P., Leaves N., Greystrom J., Coppola M., Manjunath S.,
RA Russell E., Smith M., Strachan G., Tofts C., Boal E., Cobley V.,
RA Hunter G., Kimberley C., Thomas D., Cave-Berry L., Weston P.,
RA Bochevsky M.R.M.;
RL Submitted (FBI-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL670999; CAD3017.1; -
DR GO; GO:0000785; C:Chromatin; IEA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.

DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000335; Hgcmobility_12.
DR InterPro; IPR000910; HMG_12_box.
DR Pfam; PF00505; HMG_box_2.
DR PRINTS; PR00886; HIGHMOBILITY12.
DR SMART; SM00398; HMG; 2.
DR PROSITE; PS00353; HMG_BOX_1; 1.
DR PROSITE; PS0118; HMG_BOX_2; 2.
SQ SEQUENCE 208 AA; 23811 MW; 4DBC87B9516D7B52 CRC64;

Query Match 2.2%; Score 9; DB 11; Length 208;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 293 EEDEDEDED 301
Db 191 EEDEDEDED 199

RESULT 7

O9NQJ4 PRELIMINARY; PRT; 211 AA.
AC O9NQJ4;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, last annotation update)
DE D579P20.1 (High-mobility group (Nonhistone chromosomal) protein
DE 1-like 1).
GN HMG111.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;
RN [1]
RE SEQUENCE FROM N.A.
RC Howden P.;
RL Submitted (FBI-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL160176; CAB92731.1; -
DR HSSP; P07155; 1HMF.
DR GO; GO:000785; C:chromatin; IEA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000135; Highmobility_12.
DR InterPro; IPR000910; HMG_12_box.
DR Pfam; PF00505; HMG_box_2.
DR PRINTS; PR00886; HIGHMOBILITY12.
DR SMART; SM00398; HMG; 2.
DR PROSITE; PS0118; HMG_BOX_2; 2.
SQ SEQUENCE 211 AA; 24238 MW; D2623FDE00FA8355 CRC64;

Query Match 2.2%; Score 9; DB 4; Length 211;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 293 EEDEDEDED 301
Db 188 EEDEDEDED 196

RESULT 8

O86612 PRELIMINARY; PRT; 215 AA.
AC O86612;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, last annotation update)
DE High mobility group protein.
GN HMG1.
OS Spalax leucodon ehrenbergi (Ehrenberg's mole rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Spalacinae;
OC Nanospalax.

```

OX NCBI_TaxID=30637;
RN [1]
RP SEQUENCE FROM N.A.
RA Lee K.-L.D., Lum H.-K., Nevo E.;
RT "HMG1 Genes from the Mole Rat Spalax ehrenbergi.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF078818; AAC27651.1; -.
DR HSSP; P07156; INHN.
DR GO; GO:0000785; C:chromatin; IEA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000135; Highmobility_12.
DR Pfam; PF00505; HMG_box; 2.
DR PRINTS; PR00886; HIGHMOBLTY12.
DR SMART; SM00398; HMG; 2.
DR PROSITE; PSS0118; HMG_BOX 2; 2.
SQ SEQUENCE 215 AA; 24923 MW; 1C6FB6845CA1E6C8 CRC64;

Query Match 2.2%; Score 9; DB 11; Length 215;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 293 EDEDEDEED 301
Db 188 EDEDEDEED 196

RESULT 9
ID Q9QWY6 PRELIMINARY; PRT; 215 AA.
AC Q9QWY6;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE High mobility group protein.
GN HMG1.
OS Spalax leucodon ehrenbergi (Ehrenberg's mole rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Spalacinae;
OC Nannospalax.
OC NCBI_TaxID=30637;
CX [1]
RN SEQUENCE FROM N.A.
RP Lee K.-L.D., Lum H.-K., Nevo E.;
RT "HMG1 Genes from the Mole Rat Spalax ehrenbergi.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF078819; AAC27652.1; -.
DR HSSP; P07156; INHN.
DR GO; GO:0000785; C:chromatin; IEA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000135; Highmobility_12.
DR InterPro; IPR000910; HMG_12_box.
DR Pfam; PF00505; HMG_box; 2.
DR SMART; SM00398; HMG; 2.
DR PROSITE; PSS0118; HMG_BOX 2; 2.
SQ SEQUENCE 215 AA; 24822 MW; D20D659274B575B4 CRC64;

Query Match 2.2%; Score 9; DB 11; Length 215;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 293 EDEDEDEED 301
Db 188 EDEDEDEED 196

RESULT 10

```

```

O9QX40
ID Q9QX40 PRELIMINARY; PRT; 215 AA.
AC Q9QX40;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE High mobility group protein.
GN HMG1.
OS Spalax leucodon ehrenbergi (Ehrenberg's mole rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Spalacinae;
OC Nannospalax.
OC NCBI_TaxID=30637;
CX [1]
RN SEQUENCE FROM N.A.
RP Lee K.-L.D., Lum H.-K., Nevo E.;
RT "HMG1 Genes from the Mole Rat Spalax ehrenbergi.";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF078820; AAC27653.2; -.
DR HSSP; P07156; INHN.
DR GO; GO:0000785; C:chromatin; IEA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000135; Highmobility_12.
DR InterPro; IPR000910; HMG_12_box.
DR Pfam; PF00505; HMG_box; 2.
DR PRINTS; PR00886; HIGHMOBLTY12.
DR SMART; SM00398; HMG_BOX 1; 1.
DR PROSITE; PSS0118; HMG_BOX 2; 2.
DR PROSITE; PSS0118; HMG_BOX 2; 2.
SQ SEQUENCE 215 AA; 24895 MW; 599FE8A6FDF41F17 CRC64;

Query Match 2.2%; Score 9; DB 11; Length 215;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 293 EDEDEDEED 301
Db 188 EDEDEDEED 196

```

```

RESULT 11
ID Q8BIJ2 PRELIMINARY; PRT; 354 AA.
AC Q8BIJ2;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Weakly similar to BA393J16.3.
GN B114266.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
CX [1]
RN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J;
RC MEDLINE=22354683; PubMed=12466851;
RX The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analyses of the mouse transcriptome based on functional annotation of
RL Nature 420:563-573 (2002).
DR EMBL; AK049344; BAC33696.1; -.
DR MGD; MGI:2143362; B114266.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001909; KRAB.
DR InterPro; IPR007087; ZnF_C2H2.
DR Pfam; PF01352; KRAB; 1.
DR Pfam; PF00096; zf-C2H2; 7.

```

DR SMART: SM00349; KRAB; 1.
DR SMART: SM00355; ZNF_C2H2; 7.
DR PROSITE: PS00805; KRAB; 1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 7.
DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 7.
SQ SEQUENCE 354 AA; 41447 MW; F2C3D6A421B43B4 CRC64;

Query Match
Best Local Similarity 2.2%; Score 9; DB 11; Length 354;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 KAFFRKSHL 72
DB 220 KAFFRKSHL 228

RESULT 12
Q9S2I2 PRELIMINARY; PRT; 372 AA.
AC Q9S2I2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Nucleosome assembly protein I-like protein.
GN P20B18.220 OR A74G26110.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidops.
ON NCBI_Taxid=3702;
[1]
RP SEQUENCE FROM N.A.
RA Bryan M., Rose M., Hempel S., Entian K.-D., Hohnselt J., Mewes H.W.,
RA Mayer K.F.X., Schueler C.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Rose M., Hempel S., Entian K.-D., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carrincci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Becker J., Theologis A., Davis R.W.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carrincci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Becker J., Theologis A., Davis R.W.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN [7]
RP EMBL; AL049483; CAB39676.1; -
DR EMBL; AL161564; CAB79466.1; -
DR EMBL; AY054484; AAK96675.1; -
DR EMBL; AY053283; AAM13282.1; -
DR PIR; T04266; T04266.
GO; GO:0005634; C:nucleus; IEA.
GO; GO:0003677; F:DNA binding; IEA.
GO; GO:0006334; P:nucleosome assembly; IEA.

DR InterPro: IPR002164; NAP_family.
DR Pfam: PF00956; NAP; 1.
SQ SEQUENCE 372 AA; 42997 MW; 8AA1C7B84C0BC4D2 CRC64;

Query Match
Best Local Similarity 2.2%; Score 9; DB 10; Length 372;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 293 EDEDEDED 301
DB 321 EDEDEDED 329

RESULT 13
Q38809 PRELIMINARY; PRT; 382 AA.
AC Q38809;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Col-0 nucleosome assembly protein I-like protein (Fragment).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidops.
ON NCBI_Taxid=3702;
[1]
RP SEQUENCE FROM N.A.
RA STRAIN=Col-0; TISSUE=leaf;
RX MEDLINE=95007758; PubMed=7923358;
RA Mindinos M., Katagiri F., Yu G., Ausubel F.M.;
RT "The A. thaliana disease resistance gene RPS2 encodes a protein
RT containing a nucleotide-binding site and leucine-rich repeats."
RL Cell 78:1089-1099(1994).
DR EMBL; U12858; AAAS0234.1; -
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006334; P:nucleosome assembly; IEA.
DR InterPro: IPR002164; NAP_family.
DR Pfam: PF00956; NAP; 1.
FT NON_TER
SQ SEQUENCE 382 AA; 44064 MW; EA0CB83111C0F75E CRC64;
Query Match
Best Local Similarity 2.2%; Score 9; DB 10; Length 382;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 293 EDEDEDED 301
DB 331 EDEDEDED 339

RESULT 14
Q9DFB9 PRELIMINARY; PRT; 390 AA.
AC Q9DFB9;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Oncoprotein myc.
GN MYC.
OS Ictalurus punctatus (Channel catfish).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Siluriformes;
OC Ictaluridae; Ictalurus.
ON NCBI_Taxid=7998;
[1]
RP SEQUENCE FROM N.A.
RA Barker K.S., Bengten E., Wilson M.R., Miller N.W., Warr G.W.,
RA Clem L.W.;
RT "Catfish myc oncogene".
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF

TRANSCRIPTION FACTORS.

CC EMBL; AF283994; AAG14351.1; -.
 DR HSP: P25912; IHL0.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003700; F:transcription factor activity; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR001092; HLH_basic.
 DR InterPro; IPR003327; Myc-LZ.
 DR InterPro; IPR002418; TF_Myc.
 DR Pfam; PF00010; HLH; 1.
 DR Pfam; PF02344; Myc-LZ; 1.
 DR Pfam; PF01056; Myc N term; 1.
 DR PRINTS; PR00044; LEUZIPPAMYC.
 DR SMART; SM00353; HLH; 1.
 DR PROSITE; PS50888; HLH_2; 1.
 DR SEQUENCE 390 AA; 44695 MW; 2422D9AYDBD5D700 CRC64;

Query Match 2.2%; Score 9; DB 13; Length 390;
 Best Local Similarity 100.0%; Pred. No. 4.3;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 293 EDEDEDED 301
 |||||

Db 205 EDEDEDED 213

RESULT 15

Q86YR3 PRELIMINARY; PRT; 417 AA.
 AC Q86YR3;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Adult retina hypothetical protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Li F., Yao K.T.;
 RT "Characterization and chromosomal assignment of a novel human gene
 similar to Mus musculus adult retina hypothetical protein cDNA.";
 RL Submitted (NOV-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AY174896; AA018732.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 417 AA; 46579 MW; 9DBD37B07C14556B CRC64;

Query Match 2.2%; Score 9; DB 4; Length 417;
 Best Local Similarity 100.0%; Pred. No. 4.5;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 293 EDEDEDED 301
 |||||

Db 355 EDEDEDED 363

Search completed: July 22, 2004, 16:55:35
 Job time : 46 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OK protein - protein search, using sw model

Run on: July 22, 2004, 16:42:30 ; Search time 54 Seconds

(without alignments)
2155.733 Million cell updates/sec

Title: US-09-831-804-3
2229
Sequence: 1 MSESDETKSISSLISSSSSS.....PLVKAKMDILPNETSVISR 412

Scoring table: BLOSUM62
Gap 10.0 , Gapext 0.5

Searched: 1586107 segs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: Geneseqp29Jan04:*
2: geneseqp1980s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 2229 | 100.0 | 412 | 3 | AAy93316 |
| 2 | 2225 | 99.8 | 412 | 5 | ABp73656 |
| 3 | 445 | 20.0 | 564 | 6 | ABU25655 |
| 4 | 445 | 20.0 | 564 | 6 | ABJ26255 |
| 5 | 416.5 | 18.7 | 1212 | 4 | ABG00399 |
| 6 | 403 | 18.1 | 1230 | 4 | AAU30831 |
| 7 | 400 | 17.9 | 809 | 4 | AAAM38689 |
| 8 | 398 | 17.9 | 563 | 6 | ABU936690 |
| 9 | 395.5 | 17.7 | 365 | 3 | AAy93317 |
| 10 | 395.5 | 17.7 | 409 | 4 | AAg75181 |
| 11 | 395.5 | 17.7 | 423 | 2 | AAAR91305 |
| 12 | 395 | 17.7 | 622 | 4 | AAAT78947 |
| 13 | 395 | 17.7 | 631 | 4 | AAAT79931 |
| 14 | 395 | 17.7 | 632 | 4 | ABG18386 |
| 15 | 395 | 17.7 | 719 | 4 | ABG16953 |
| 16 | 393.5 | 17.7 | 1472 | 4 | AAU31578 |
| 17 | 393 | 17.6 | 517 | 4 | ABG03375 |
| 18 | 392.5 | 17.6 | 1050 | 7 | ADG32995 |
| 19 | 392.5 | 17.6 | 1050 | 7 | ADG32995 |
| 20 | 392 | 17.6 | 555 | 6 | ADA54763 |
| 21 | 390.5 | 17.5 | 878 | 4 | AAU29528 |
| 22 | 390.5 | 17.5 | 1520 | 4 | ABG27130 |
| 23 | 390 | 17.5 | 540 | 7 | ADB65594 |
| 24 | 390 | 17.5 | 576 | 6 | ADA54797 |
| 25 | 390 | 17.5 | 620 | 6 | ABU96725 |

| | | | | | |
|----|-------|------|-----|---|-----------|
| 26 | 388.5 | 17.4 | 501 | 4 | ABB71369 |
| 27 | 388 | 17.4 | 568 | 5 | ABB79872 |
| 28 | 387.5 | 17.4 | 872 | 7 | ADG37535 |
| 29 | 386 | 17.3 | 727 | 4 | ABG16954 |
| 30 | 385.5 | 17.3 | 474 | 6 | ABU11782 |
| 31 | 385 | 17.3 | 577 | 5 | AAE14680 |
| 32 | 385 | 17.3 | 770 | 4 | ABG10091 |
| 33 | 385 | 17.3 | 770 | 4 | ABG14920 |
| 34 | 385 | 17.3 | 803 | 4 | ABG20106 |
| 35 | 380 | 17.0 | 615 | 6 | AAE37047 |
| 36 | 380 | 17.0 | 959 | 6 | ABG01956 |
| 37 | 379 | 17.0 | 817 | 4 | AAAM40475 |
| 38 | 378 | 17.0 | 613 | 4 | AAAB95862 |
| 39 | 378 | 17.0 | 755 | 4 | AAAM40916 |
| 40 | 378 | 17.0 | 907 | 4 | AAAB94428 |
| 41 | 377 | 16.9 | 518 | 6 | ABR41469 |
| 42 | 376 | 16.9 | 553 | 7 | ADG31307 |
| 43 | 376 | 16.9 | 582 | 6 | ABR41407 |
| 44 | 376 | 16.9 | 675 | 5 | ABP55423 |
| 45 | 376 | 16.9 | 678 | 7 | ADG31847 |

ALIGNMENTS

RESULT 1
AAy93316
ID AAy93316 standard; protein; 412 AA.

AC AAy93316;
DT 04-SEP-2000 (first entry)

DE A transcription factor designated CATP11A.
KW Transcription factor; CATP11A; DNA-binding protein;
KM ribosomal RNA 5S gene; fungal infection.
XX

OS Candida albicans.

PH Key Location/Qualifiers

FT Misc-difference 193 /note= "Ser encoded by CTG"

FT FT /note= "Ser encoded by CTG"

FT FT /note= "Ser encoded by CTG"

FT FT /note= "Ser encoded by CTG"

FT FT /note= "Ser encoded by CTG"

FT FT /note= "Ser encoded by CTG"

FT FT /note= "Ser encoded by CTG"

FT FT /note= "Ser encoded by CTG"

FT FT /note= "Ser encoded by CTG"

FT FT /note= "Ser encoded by CTG"

FT FT /note= "Ser encoded by CTG"

FT FT /note= "Ser encoded by CTG"

FT FT /note= "Ser encoded by CTG"

FT FT /note= "Ser encoded by CTG"

FT FT /note= "Ser encoded by CTG"

FT FT /note= "Ser encoded by CTG"

FT FT /note= "Ser encoded by CTG"

FT FT /note= "Ser encoded by CTG"

FT FT /note= "Ser encoded by CTG"

FT FT /note= "Ser encoded by CTG"

FT FT /note= "Ser encoded by CTG"

FT FT /note= "Ser encoded by CTG"

FT FT /note= "Ser encoded by CTG"

FT FT /note= "Ser encoded by CTG"

FT FT /note= "Ser encoded by CTG"

FT FT /note= "Ser encoded by CTG"

FT FT /note= "Ser encoded by CTG"

FT FT /note= "Ser encoded by CTG"

FT FT /note= "Ser encoded by CTG"

FT FT /note= "Ser encoded by CTG"

FT FT /note= "Ser encoded by CTG"

FT FT /note= "Ser encoded by CTG"

FT FT /note= "Ser encoded by CTG"

FT FT /note= "Ser encoded by CTG"

Claim 12, Page 35-36; 45pp; French.
The present sequence represents a Candida albicans transcription factor, designated CATP11A. The polypeptide is a DNA-binding protein, which is involved in initiating transcription of the ribosomal RNA 5S gene. The polynucleotide is used to screen for its specific inhibitors, potentially useful as antimycotic agents, to raise an antibody response that is protective against fungal infection and to raise antibodies. Such antibodies, as well as the polypeptides and polynucleotides are used in

CC compositions for diagnosing and treating fungal infections, e.g. by
CC detecting polymorphisms and mutations
XX
SQ Sequence 412 AA:
Query Match 100.0%; Score 2229; DB 3; Length 412;
Best Local Similarity 100.0%; Pred. No. 1e-167;
Matches 412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSESDETKSISSIISSSSSRPKKYYICTYEGCDKAYNRPSELBOHLRTHSNDRPYKCTVD 60
Db 1 MSESDETKSISSIISSSSSRPKKYYICTYEGCDKAYNRPSELBOHLRTHSNDRPYKCTVD 60
QY 61 DCDKAFPRKSHLETHIVSHSEKPPHCSCGVGKYNRQHLKRHEITHTKSPKCTFENCQE 120
Db 61 DCDKAFPRKSHLETHIVSHSEKPPHCSCGVGKYNRQHLKRHEITHTKSPKCTFENCQE 120
QY 121 AFYKHQSLRHHTLSVHEKTLTCQCKNVFTRPSKLAQHKLKHGGSPAYQCDHPGCFKMF 180
Db 121 AFYKHQSLRHHTLSVHEKTLTCQCKNVFTRPSKLAQHKLKHGGSPAYQCDHPGCFKMF 180
QY 181 QTWSVLQPHIKOSHPRKLCPCGKGVGKGLSSHMLSHDSTWIKITWCDYCDVGKFAK 240
Db 181 QTWSVLQPHIKOSHPRKLCPCGKGVGKGLSSHMLSHDSTWIKITWCDYCDVGKFAK 240
QY 241 KNEIVEHYNIFHDGNIPDDLKETEVEKKLENLDQSKLANLHLETEKLVKVEDEDEE 300
Db 241 KNEIVEHYNIFHDGNIPDDLKETEVEKKLENLDQSKLANLHLETEKLVKVEDEDEE 300
QY 301 DSIDEKRSVDSMSAQRSIKSFASLEGSKSVSKLISNGKKINCPKNNCDRMFSREY 360
Db 301 DSIDEKRSVDSMSAQRSIKSFASLEGSKSVSKLISNGKKINCPKNNCDRMFSREY 360
QY 361 DLRHLLKMHDDNLQRIESFLNSIEKEETPEGEPLVKKARMDLLPNETSVISR 412
Db 361 DLRHLLKMHDDNLQRIESFLNSIEKEETPEGEPLVKKARMDLLPNETSVISR 412
RESULT 2
ABP73656
ID ABP73656 standard; protein; 412 AA.
XX
AC ABP73656;
XX
DT 30-JAN-2003 (first entry)
XX
DE Candida albicans essential protein SEQ ID NO 7493.
XX
XX Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;
KW signal transduction; DNA replication; cell division; growth;
KW proliferation; Candida albicans; fungicide; antifungal.
XX
OS Candida albicans.
XX
PN WO200253728-A2.
XX
PD 11-JUL-2002.
XX
PF 26-DEC-2001; 2001WO-US049486.
XX
PR 29-DEC-2000; 2000US-0259128P.
PR 20-FEB-2001; 2001US-00792024.
PR 22-AUG-2001; 2001US-0314050P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
PI Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL,
XX
DR WPI; 2002-566694/60.
XX
DR N-PSDB; AB232206.
XX
PT Constructing strains for identifying gene products as effective targets
PT for therapeutic intervention, by inactivating in the strain one allele of

PT a gene and placing other allele of the gene under conditional expression.
XX
XX Claim 44; SEQ ID NO 7493; 167pp + Sequence Listing; English.
PS
XX The invention relates to constructing (M1) a strain of diploid fungal
CC cells in which both alleles of a gene are modified, comprising modifying
CC one allele by insertion or replacement by a cassette having an
CC expressible selectable marker and modifying other allele by
CC recombination, of a promoter replacement fragment with a heterologous
CC promoter, so that expression of the second allele is regulated by the
CC promoter. (M1) is useful for constructing a strain of diploid fungal
CC cells in which both alleles of a gene are modified. The diploid fungal
CC cells having both alleles modified are useful for identifying a gene that
CC is essential to the survival or growth of a fungus, a gene that
CC contributes to the virulence and/or pathogenicity of a fungus, a gene
CC that contributes to the resistance of a diploid fungus to an antifungal
CC agent, an antifungal agent that inhibits the growth of a mammalian
CC agent, and for identifying a therapeutic agent for treatment of a mammalian
CC disease. (M1) is useful for identifying a compound which modulates the
CC activity of a gene product, preferably enzymatic activity, carbon
CC compound catabolism, biosynthesis, transporter, transcriptional,
CC translational, signal transduction, DNA replication and cell division
CC activity. The method is useful for identifying a compound having the
CC ability to inhibit growth or proliferation of C. albicans cells and for
CC treating infection by C. albicans. The present sequence is that of an
CC essential Candida albicans protein used in the method of the invention.
CC Note: The sequence data for this patent is not represented in the printed
CC specification but is based on sequence information supplied to Derwent by
CC the European Patent Office
XX
SQ Sequence 412 AA:
Query Match 99.8%; Score 2225; DB 5; Length 412;
Best Local Similarity 99.8%; Pred. No. 2.1e-167;
Matches 411; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSESDETKSISSIISSSSSRPKKYYICTYEGCDKAYNRPSELBOHLRTHSNDRPYKCTVD 60
Db 1 MSESDETKSISSIISSSSSRPKKYYICTYEGCDKAYNRPSELBOHLRTHSNDRPYKCTVD 60
QY 61 DCDKAFPRKSHLETHIVSHSEKPPHCSCGVGKYNRQHLKRHEITHTKSPKCTFENCQE 120
Db 61 DCDKAFPRKSHLETHIVSHSEKPPHCSCGVGKYNRQHLKRHEITHTKSPKCTFENCQE 120
QY 121 AFYKHQSLRHHTLSVHEKTLTCQCKNVFTRPSKLAQHKLKHGGSPAYQCDHPGCFKMF 180
Db 121 AFYKHQSLRHHTLSVHEKTLTCQCKNVFTRPSKLAQHKLKHGGSPAYQCDHPGCFKMF 180
QY 181 QTWSVLQPHIKOSHPRKLCPCGKGVGKGLSSHMLSHDSTWIKITWCDYCDVGKFAK 240
Db 181 QTWSVLQPHIKOSHPRKLCPCGKGVGKGLSSHMLSHDSTWIKITWCDYCDVGKFAK 240
QY 241 KNEIVEHYNIFHDGNIPDDLKETEVEKKLENLDQSKLANLHLETEKLVKVEDEDEE 300
Db 241 KNEIVEHYNIFHDGNIPDDLKETEVEKKLENLDQSKLANLHLETEKLVKVEDEDEE 300
QY 301 DSIDEKRSVDSMSAQRSIKSFASLEGSKSVSKLISNGKKINCPKNNCDRMFSREY 360
Db 301 DSIDEKRSVDSMSAQRSIKSFASLEGSKSVSKLISNGKKINCPKNNCDRMFSREY 360
QY 361 DLRHLLKMHDDNLQRIESFLNSIEKEETPEGEPLVKKARMDLLPNETSVISR 412
Db 361 DLRHLLKMHDDNLQRIESFLNSIEKEETPEGEPLVKKARMDLLPNETSVISR 412
RESULT 3
ABJ25655
ID ABJ25655 standard; protein; 564 AA.
XX
AC ABJ25655;
XX
DT 16-APR-2003 (first entry)
XX

| | | |
|----|---|---|
| Dc | | Aspergillus fumigatus essential gene protein #313. |
| Xw | Kw | Fungicide; cytosstatic; essential gene; Aspergillus fumigatus; infection; |
| Xw | Kw | cancer; contamination; biofilm; antibody; immune response. |
| Ox | Ss | Aspergillus fumigatus. |
| Pn | | W0200286090-A2. |
| Pd | | 31-OCT-2002. |
| Pf | | 23-APR-2002; 2002WC-US013142. |
| Px | | 23-APR-2001; 2001US-0285697P. |
| Pr | | 27-APR-2001; 2001US-0287066P. |
| Pr | | 05-JUN-2001; 2001US-0295890P. |
| Pr | | 09-JUL-2001; 2001US-0303899P. |
| Px | | 31-AUG-2001; 2001US-0316362P. |
| Pa | (ELIT-) | ELITRA PHARM INC. |
| Pi | Jiang B, | Tshkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM; |
| Pt | | WPI; 2003-093124/08. |
| Dr | | New purified or isolated nucleic acids of essential genes of Aspergillus |
| Xx | | fumigatus, useful for treating or preventing infections by A. fumigatus, |
| Xx | | or for treating a non-infectious disease in a subject e.g. cancer. |
| Ps | | Disclosure; Page; 175pp; English. |
| Cc | | The invention relates to novel purified or isolated nucleic acids of |
| Cc | | essential genes of Aspergillus fumigatus. The isolated nucleic acids of |
| Cc | | the invention are used to treat or prevent infections by a pathogenic |
| Cc | | organism such as A. fumigatus, to treat a non-infectious disease in a |
| Cc | | subject (e.g. cancer), to prevent or contain contamination of an object |
| Cc | | by A. fumigatus, or to prevent or inhibit formation on a surface of a |
| Cc | | biofilm comprising A. fumigatus. The polynucleotides are useful for |
| Cc | | expressing recombinant protein for characterisation, screening or |
| Cc | | therapeutic use, as markers for host tissues in which the pathogenic |
| Cc | | organisms invade or reside, for comparing with the DNA sequence of A. |
| Cc | | fumigatus to identify duplicated genes or paralogues having the same or |
| Cc | | similar biochemical activity and/or function, for comparing with DNA |
| Cc | | sequences of other related or distant pathogenic organisms to identify |
| Cc | | potential orthologous essential or virulence genes, for selecting and |
| Cc | | making oligomers for attachment to a nucleic acid array for examination |
| Cc | | of expression patterns, for raising anti-protein antibodies, as an |
| Cc | | antigen to raise anti-DNA antibodies or to elicit another immune |
| Cc | | response, and for identifying polynucleotides encoding the other protein |
| Cc | | with which binding occurs or to identify inhibitors of the binding |
| Cc | | interaction. The polypeptides may be used to raise antibodies or to |
| Cc | | elicit immune response, as a reagent in assays designed to quantitatively |
| Cc | | determine levels of the protein in biological fluids, as a marker for |
| Cc | | host tissues in which pathogenic organism invade or reside, and to |
| Cc | | isolate correlative receptors or ligands in the case of virulence |
| Cc | | factors. This sequence represents a protein of one of the essential genes |
| Cc | | of Aspergillus fumigatus of the invention |
| Sq | | Sequence 564 AA; |
| Qy | Query Match | 20.0%; Score 445; DB 6; Length 564; |
| Dc | Best Local Similarity | 30.3%; Pred. No. 1.8e-26; |
| Mt | Matches 125; Conservative | 62; Mismatches 164; Indels 62; Gaps 21 |
| 14 | ISSSSSP--KXICTYEGCDKAYNPISLLEOHRTSNDPRPYCYTDDCCDAFFRS | 70 |
| 71 | ISTTSAKPSSELTCKTRCPDGGCTKAFNRPARIOEHRSHNNERIFECTEECDKFTLRAS | 130 |
| 71 | HLETHIVS-HSEKKPFHCS--VCGGVNSRQLKHKEITH--TKSPKT-FENCCEAFYK | 124 |
| Db | HLNHIKSAHGVDYDCDRPGCGSFYTGSILRHIAHDGRDKRCIEYPPCMTTFRK | 190 |
| Yz | HQSIRHHLSVH--EKYLTC-----KQCNKVFTFRBSKLQHKLKTHGSGPAYQCDH- | 173 |

[illegible]

| | | | |
|----------|----------|---|------|
| 0y | 311 | RSDMSAQRISFTALBESKYSKISNSGKINCPCNNCDMFREYDLRRHLMK | 369 |
| Db | 1152 | REKPYKKECGKSPNRSSTFTK--HKVITHGVKLYKC--EECGKSPFWSALTRHKH | 1200 |
| RESULT 7 | | | |
| ID | AAM38689 | standard; protein; 809 AA. | |
| XX | AC | AAM38689; | |
| XX | DT | 22-OCT-2001 (first entry) | |
| DE | XX | Human polypeptide SEQ ID NO 1834. | |
| XX | XX | | |
| KW | KW | Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer; | |
| KW | KW | peripheral nervous system; neuropathy; central nervous system; CNS; | |
| KW | KW | Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; | |
| KW | KW | amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic; | |
| KW | KW | chemokinetic; thrombolytic; drug screening; arthritis; inflammation; | |
| XX | XX | leukaemia. | |
| OS | XX | Homo sapiens. | |
| PN | XX | MO200153312-A1. | |
| PD | XX | 26-UTL-2001. | |
| XX | XX | | |
| PF | XX | 26-DEC-2000; 2000WO-US034263. | |
| XX | XX | | |
| PR | XX | 23-DEC-1999; 99US-00471275. | |
| PR | XX | 21-JAN-2000; 2000US-00486725. | |
| PR | XX | 25-APR-2000; 2000US-00552317. | |
| PR | XX | 20-JUN-2000; 2000US-00598042. | |
| PR | XX | 19-JUL-2000; 2000US-00620312. | |
| PR | XX | 03-AUG-2000; 2000US-00653450. | |
| PR | XX | 14-SEP-2000; 2000US-00662191. | |
| PR | XX | 19-OCT-2000; 2000US-00693036. | |
| PR | XX | 29-NOV-2000; 2000US-00727344. | |
| PA | XX | (HYSE-) HYSEQ INC. | |
| PI | XX | | |
| PI | XX | Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D; | |
| PI | XX | Wang J, Wang Z, Weinman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA; | |
| PI | XX | Zhou P, Gooddich R, Dimaac RT; | |
| DR | XX | WPI; 2001-442253/47. | |
| DR | XX | N-PSDB; AAI57845. | |
| PT | XX | | |
| PT | XX | Novel nucleic acids and polypeptides, useful for treating disorders such | |
| PS | XX | as central nervous system injuries. | |
| XX | XX | | |
| XX | XX | Example 3; SEQ ID NO 1834; 10078pp; English. | |
| CC | CC | | |
| CC | CC | The invention relates to human nucleic acids (AAI57798-AAI61369) and the | |
| CC | CC | encoded polypeptides (AAM38642-AA42213) with nootropic, | |
| CC | CC | immunosuppressant and cytostatic activity. The polynucleotides are useful | |
| CC | CC | in gene therapy. A composition containing a polypeptide or polynucleotide | |
| CC | CC | of the invention may be used to treat diseases of the peripheral nervous | |
| CC | CC | system, such as peripheral nervous injuries, peripheral neuropathy and | |
| CC | CC | localised neuropathies and central nervous system diseases, such as | |
| CC | CC | Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic | |
| CC | CC | lateral sclerosis, and Shy-Drager Syndrome. Other uses include the | |
| CC | CC | utilitation of the activities such as: Immune system suppression, | |
| CC | CC | Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic | |
| CC | CC | and thrombolytic activity, cancer diagnosis and therapy, drug screening, | |
| CC | CC | assays for receptor activity, arthritis and inflammation, leukaemias and | |
| CC | CC | C.N.S disorders. Note: The sequence data for this patent did not form | |
| XX | XX | part of the printed specification | |
| XX | XX | Sequence 809 AA; | |

Best Local Similarity 31.9%; Pred. No. 1.1e-22; Indels 34; Gaps 19;
Matches 121; Conservative 53; Mismatches 171;
QY 2 SESEDEKSISSIISSSSSS--SRPKKYYCYEGCDKAYNRPISLEQHLRTHSNDPRPKCTV 59
Db 373 TEGGAFSSSSNLTTHKKIHTGKPKYKC--EECGKAFKMSSTLTKHLTHGKPKYKC-- 428
QY 60 DDCDKAFPKRSHLETHIVSHSEKKPHGCVGKGVNSRQHLKHEITHT--KSFKCTPEN 117
Db 429 EECGKAFNRPSTLTKNRIHTGKPKYCEVCGKAFQFSNLTTHKRIHTGKPKYKC--EE 486
QY 118 CGGAFKQHSI--RHHILSVHEKTLTCKQCNKVFTRPSKLAQHLKHHGSPAYQCDHPG 176
Db 487 CGGAFRRSSNLTTHKKIHTGKPKYCBEGCGKAFKMSSTLTKHLTHGKPKYKC--C 543
QY 177 FKNFQTSVYLOPH--IKQSHPKLKCPKCGKGVGKGLSSHMLSHDSTMIKTWDCYD 234
Db 544 GKAFNRPSTLTTHKKIHTGKPKYCEVCGKAFQFSNLTTHKKIHTGKPKYKC--EE 599
QY 235 VGR-FAKKNELVHYNIFHDGNIPDILLKTEYKLENNLDQSKLNNHELETKLQVE 293
Db 600 -GKAFQSSNLTTHKKI-HTGKRP--YKCECGKAFNRPSTLTTHKKIHTGKPKYCB 653
QY 294 E-DEBDEDSLDEKSDVRSDS--MSAQRSIKSFPLSLGSKSVSKLISNGKKINCPKN 350
Db 654 ECGKAFKMSSTLTTHKKIHTGKPKYCBEGCGKAF--KLSSTLSHTKLIHTGKPKYCEK- 710
QY 351 NCDRMPREYDLRRHLMKH 369
Db 711 -CGKAFRRPSNLTTHKKIHTGKPKYCBEGCGKAF--KLSSTLSHTKLIHTGKPKYCEK- 728
RESULT 8
ABU96690 standard; protein, 563 AA.
AC ABU96690;
XX 25-JUL-2003 (first entry)
DT
XX Human nucleic acid-associated protein (NAAP) #19.
DE
XX Human, nucleic acid-associated protein; cytosolic; antiarteriosclerotic;
KM anticovulstant; nootropic; neuroprotective; cerebroprotective; anti-HIV;
KM antifibrotic; antiinflammatory; chymotrypsin; gene therapy;
KM cell proliferative disorder; cancer; atherosclerosis;
KM neurological disorder; epilepsy; Huntington's disease; stroke;
KM immune disorder; inflammatory disorder; AIDS; allergy;
KM developmental disorder; hypothyroidism; Cushing's syndrome; infection;
KM protein-protein interaction; drug-target interaction;
KM gene expression profile.
XX Homo sapiens.
OS
XX WO2003023003-A2.
EN
XX 20-MAR-2003.
PD
XX 05-SEP-2002; 2002WO-US028540.
PF
XX 07-SEP-2001; 2001US-0317792P.
PR 07-SEP-2001; 2001US-0317912P.
PR 14-SEP-2001; 2001US-032270P.
PR 21-SEP-2001; 2001US-0324040P.
PR 28-SEP-2001; 2001US-0326732P.
PR 19-OCT-2001; 2001US-0346716P.
PR 25-JAN-2002; 2002US-0351749P.
PR 22-FEB-2002; 2002US-0359498P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX Tang YF, Jackson JL, Griffin JA, Elliott VS, Forsythe IU;
PI Becha SD, Richardson TW, Lee EA, Sprague WM, Emerling BM;

PI Thangavelu K, Warren BA, Tran UK, Yue H, Xu Y, Yue H, Li JX;
PI Hafalia AU, Sanjanwala B, Marguis JP, Gorvad AE, Lee SY, Ison CH,
PI Raughn MR, Chawla NK, Nguyen DB, Swarnakar A, Zebairjahan Y, Shah P;
PI Thornton M, Yao MG, Khan FA, Gandhi AR, Yang U, Kahle AE;
PI Burford N, Ramkumar J;
XX
XX MPI: 2003-313243/30.
DR N-PDB: ACA98938.
XX
XX New human nucleic acid associated proteine (NAAP), useful for diagnosing,
PT treating and preventing diseases or conditions associated with the
PT aberrant NAAP expression e.g. cancer, AIDS, atherosclerosis, epilepsy, or
PT infections.
PS Claim 1; Page 255-257; 345pp; English.
XX
XX The invention describes a novel human isolated nucleic acid-associated
CC polypeptide (NAAP). The polypeptides and polynucleotides are useful in
CC diagnosing, treating and preventing diseases or conditions associated
CC with the decreased expression or overexpression of NAAP, such as cell
CC proliferative (e.g. cancer, atherosclerosis), neurological (e.g.
CC epilepsy, Huntington's disease, stroke), immune/inflammatory (e.g. AIDS,
CC allergies) and developmental (e.g. Hypothyroidism, Cushing's syndrome)
CC disorders, or infections. These are also useful in assessing the effects
CC of exogenous compounds on the expression of nucleic acid and amino acid
CC sequences of NAAP. The NAAP or its fragments are useful in screening
CC compounds for effectiveness as agonist or antagonist of the polypeptides,
CC or in altering the expression of the target polynucleotide and compounds
CC that specifically bind to or modulate the activity of the polypeptide.
CC The microarray is useful in monitoring or measuring protein-protein
CC interactions, drug-target interactions, and gene expression profiles.
CC This is the amino acid sequence of a novel human nucleic acid-associated
CC protein (NAAP)
XX
SQ Sequence 563 AA:
Query Match 17.9%; Score 398; DB 6; Length 563;
Best Local Similarity 30.2%; Pred. No. 9.3e-23;
Matches 121; Conservative 48; Mismatches 167; Indels 64; Gaps 18;
QY 20 SRPKKYYCYEGCDKAYNRPISLEQHLRTHSNDPRPKCTVDDCKAFPKRSHLETHIVSH 79
Db 196 SGEKPKYK--KECGKAFNRPSTLTTHKKIHTGKPKYKC--EECGKAFNRPSTLTTHKKIHT 251
QY 80 SEKKPHGCVGKGVNSRQHLKHEITHT--KSFKCTPENCGEAFYKQSL-RHHILSVH 136
Db 252 TGGKPRYCEBEGKAFNRPSTLTTHKKIHTGKPKYKC--EECGKAFNRPSTLTTHKKIHTG 309
QY 137 EKTLLTQCNKVFTRPSKLAQHLKHHGSPAYQCDHGCCKNRPQTSVLOPH--IKQSH 194
Db 310 EKPYPKEBEGKAFNRPSTLTTHKKIHTGKPKYCEB--CGKAFNRPSTLTTHKKIHTG 366
QY 195 PKLNCPPKCGKGVGKGLSSHMLSHDSTMIKIWTCDYCDVGFKNKNELVHYNIFHDG 254
Db 367 KPYKCEBEGKAFNRPSTLTTHKKIHTGKPKYCEB--CGKAFNRPSTLTTHKKIHTG 421
QY 255 NIPDILLKTEYKLENNLDQSKLNNHELETKLQVEBDEBDEDSLDEKSDVRSDS 314
Db 422 EKP--YKCECGKAFNRPSTLTTHKKIHTGKPKYCEB--CGKAFNRPSTLTTHKKIHTG 467
QY 315 MSAQRST-----KSFPLSLGSKSVSKLISNGKKINCPKNCDRMPREYD 361
Db 468 LSKHVIHTGKPKYCEBEGKAFNRPSTLTTHKKIHTGKPKYCEB--CGKAFNRPSTLTTH 523
QY 362 LRRHLKWHD-----DNIDRLESFINSIEKEE 387
Db 524 LRRHVIHTGKPKYCEBEGKAFNRPSTLTTHKKIHTGKPKYCEB--CGKAFNRPSTLTTH 563
RESULT 9
AAV93317
ID AAV93317 standard; protein, 365 AA.
XX

CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC (AAK52582) and 3666 (AAH80020) are omitted as the relevant pages from the
CC sequence listing were missing at the time of publication

XX SQ Sequence 622 AA;

Query Match 17.7%; Score 395; DB 4; Length 622;
Best Local Similarity 29.7%; Pred. No. 1.8e-22;
Matches 123; Conservative 42; Mismatches 157; Indels 92; Gaps 20;

```
QY 23 KKYICTYEGCDKAVNRPSLLEOHILRTHSNDREPYKTVDDCKAFPKSHLETHIVSHSEK 82
D 201 KPYIC--ECCGKAFKXSSALNTWTKRIHTGEKPYKC--DKCDKAFIASSTLSKHEIHTGK 256
QY 83 KPFHCVCGKGVNSROHLKRHEIHT--KSPKCFPENOGEAFYHQSLRHH----- 131
D 257 KPYCECGKAFNQSSTLTGKRIHTGEKPYKC--ECCGKAFNQSSTLTGKRIHTGEKPY 314
QY 132 -----ILSVH-----EKLITCKQCNKVFTPRPSKLAQHLKH----- 162
D 315 YVCECCGKAFKXSRILTTTKRIHTGEKPYKCNKCGKAFIASSTLSRHEFIHMGKHYKCE 374
QY 163 -----HGSPPAYQCDHPGCFKNQUTWSVLOPHIKOSH--PKLKCPK 201
D 375 ECGKAFIWSVLTTRHKRVHTGEKPYKCE--CGKAFKXSSSTLSH--KRSHTEKPYKCE 431
QY 202 CGKGVCGKGLSSHMLSHDSTMIKIWTCDYCDVGK--FAKNELVEHYNI FHDGNI PDDL 260
D 432 CGKAFVASSSTLSKHEIHT--TGKRPYKCEB--GKAFNQSSTLTGKRIHTGEKPY 482
QY 261 LKETEYKYLENLDDGSKLNNLHLELT--EKLVEDEDEDEDS--LDEKRSDVRSDSM 315
D 483 YKCECGKAFN--QSSSLTKHKRIHTGEKPYKCECGKAFNQSSTLTGKRIHTGEKPY 539
QY 316 SAQRISKSFYASLEGSKSVSGLISNSGSKTNCPRKNCDFRFRDYDLRHLKWH 369
D 540 KCECGKAF--HLSTHLTTHKILHTGEKPYRC--RECGKAFNHSATLSHKKIH 589
```

RESULT 13
AAH79931
ID AAM79931 standard; protein; 631 AA.

XX AC AAM79931;

XX DT 06-NOV-2001 (first entry)

XX DE Human protein SEQ ID NO 3577.

XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.

OS Homo sapiens.

XX FN WO200157190-A2.

XX PD 09-AUG-2001.

XX PF 05-FEB-2001; 2001WO-US004098.

XX PR 03-FEB-2000; 2000US-00496914.
PR 27-APR-2000; 2000US-00560875.
PR 20-JUN-2000; 2000US-00598075.
PR 19-JUL-2000; 2000US-00620335.
PR 01-SEP-2000; 2000US-00654936.
PR 15-SEP-2000; 2000US-00663561.
PR 20-OCT-2000; 2000US-00693325.
PR 30-NOV-2000; 2000US-00728422.

XX (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Dirmacac RT, Asundi V, Zhou P, Xu C, Cao Y,
PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX WPI; 2001-476283/51.

DR N-PSDB; AAK53064.

XX Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT in diagnosis and gene therapy.

XX Claim 20; Page 388; 6221pp; English.

CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAH80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, immunomodulatory activity
CC activity, tissue growth factor activity, haematopoiesis regulation
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC (AAK52582) and 3666 (AAH80020) are omitted as the relevant pages from the
CC sequence listing were missing at the time of publication

XX SQ Sequence 631 AA;

Query Match 17.7%; Score 395; DB 4; Length 631;
Best Local Similarity 29.7%; Pred. No. 1.9e-22;
Matches 123; Conservative 42; Mismatches 157; Indels 92; Gaps 20;

```
QY 23 KKYICTYEGCDKAVNRPSLLEOHILRTHSNDREPYKTVDDCKAFPKSHLETHIVSHSEK 82
D 210 KPYIC--ECCGKAFKXSSALNTWTKRIHTGEKPYKC--DKCDKAFIASSTLSKHEIHTGK 265
QY 83 KPFHCVCGKGVNSROHLKRHEIHT--KSPKCFPENOGEAFYHQSLRHH----- 131
D 266 KPYCECGKAFNQSSTLTGKRIHTGEKPYKC--ECCGKAFNQSSTLTGKRIHTGEKPY 323
QY 132 -----ILSVH-----EKLITCKQCNKVFTPRPSKLAQHLKH----- 162
D 324 YVCECGKAFKXSRILTTTKRIHTGEKPYKCNKCGKAFIASSTLSRHEFIHMGKHYKCE 383
QY 163 -----HGSPPAYQCDHPGCFKNQUTWSVLOPHIKOSH--PKLKCPK 201
D 384 ECGKAFIWSVLTTRHKRVHTGEKPYKCEB--GKAFKXSSSTLSH--KRSHTEKPYKCE 440
QY 202 CGKGVCGKGLSSHMLSHDSTMIKIWTCDYCDVGK--FAKNELVEHYNI FHDGNI PDDL 260
D 441 CGKAFVASSSTLSKHEIHT--TGKRPYKCEB--GKAFNQSSTLTGKRIHTGEKPY 491
QY 261 LKETEYKYLENLDDGSKLNNLHLELT--EKLVEDEDEDEDS--LDEKRSDVRSDSM 315
D 492 YKCECGKAFN--QSSSLTKHKRIHTGEKPYKCECGKAFNQSSTLTGKRIHTGEKPY 548
QY 316 SAQRISKSFYASLEGSKSVSGLISNSGSKTNCPRKNCDFRFRDYDLRHLKWH 369
D 549 KCECGKAF--HLSTHLTTHKILHTGEKPYRC--RECGKAFNHSATLSHKKIH 598
```

RESULT 14

ABG18386
ID ABG18386 standard; protein; 632 AA.

XX AC ABG18386;

XX DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #18377.

| | |
|----|---|
| XX | Human; chromosome mapping; gene mapping; gene therapy; forensic; |
| KM | food supplement; medical imaging; diagnostic; genetic disorder. |
| XX | |
| OS | Homo sapiens. |
| XX | |
| PN | WO200175067-A2. |
| XX | |
| PD | 11-OCT-2001. |
| XX | |
| PF | 30-MAR-2001; 2001WO-US0008631. |
| XX | |
| XX | 31-MAR-2000; 2000US-00540217. |
| PR | 23-AUG-2000; 2000US-00649167. |
| XX | |
| PA | (HYSE-) HYSEQ INC. |
| XX | |
| PI | Dymanac RT, Liu C, Tang YT; |
| XX | |
| DR | WPI; 2001-639362/73. |
| DR | N-PSDB; AAS82573. |
| XX | |
| PT | New isolated polynucleotide and encoded polypeptides, useful in |
| PT | diagnostics, forensics, gene mapping, identification of mutations |
| PT | responsible for genetic disorders or other traits and to assess |
| PT | biodiversity. |

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful for treating disorders of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at http://wipo.int/pub/publ/pct_sequences

| | |
|-----------------------|----------|
| Sequence | 632 AA; |
| Query Match | 17.7%; |
| Best Local Similarity | 29.7%; |
| Matches | 123; |
| Conservative | 42; |
| Mismatches | 157; |
| Indels | 92; |
| Gaps | 20; |
| Score | 395; |
| DB | 4; |
| Length | 632; |
| Pred. No. | 1.9e-22; |

| | | | |
|----|-----|---|-----|
| Qy | 23 | KXICITVEGDDKAVNPSLIEOHLRTHSNDPRKXCVWDODCKAFEFKSHLEHIVASHSK | 82 |
| Db | 211 | KPITC--EECGKAFKXSSALNTHKRLHTGKPKC--DKODKAFIASSTLSKHEIHTGX | 266 |
| Qy | 83 | KPHCSVCCKGVNSROHLKHEIHTT--KSFKCTPENCQAFYHOSLRNH----- | 131 |
| Db | 267 | KPKCECGKAFNQSSTLTGKKLHNGEKPKC--EECGKAFNQSSTLTGKKLHNGEKPK | 324 |
| Qy | 132 | -----IISVH-----EKTLLTCKQNNVTRPSTLAQHLKH----- | 162 |
| Db | 325 | YVCECGKAFKXSNILTHKRLHTGKEPKYKNNCGKAFIASSTLSRHEITMGKPKCE | 384 |
| Qy | 163 | -----HGSPAYOCDHDGCFENFQMTVSLQFHLKOSH--PKLKCPK | 201 |
| Db | 385 | ECGKAFIWSVLTDEHKVHTGKEPKCEE--CGKAFIWSSTLSH--KRSHITGKPKPKCEB | 441 |

| | | | |
|----|-----|--|-----|
| QY | 202 | CGKGVGKKGKSSKHLSPDSDSMIMICWDYCDVGK-FAKKKEIVENYVNI PFHGNIPDDL | 260 |
| Db | 442 | CGKAFVASTLSKHEI IH--TGKPPYKCEEC-GKAFQSSSLTHKKI-HGKEP---- | 432 |
| QY | 261 | LKTEPVKLEMLDDGSKLNNHELET--EKLVEDEDEDEEDS--LDEKSDVRSDSM | 315 |
| Db | 493 | YKCECGKAFN--OSSSLTHKKIKHTGKPPYKCECGKAFNSSLTHKKIKHTPEXPY | 549 |
| QY | 316 | SAORSIKSFSTASIEGSKSVSKLISNGKKINCPKNNCDMFSBEYDLRRHLKMH | 369 |
| Db | 550 | KCEBGKAF--HLSHTLHTHKLHNGEKYRC-REGCAFNHSAVTLSSHKHT | 599 |

RESULT 15
ABG16953
ID ABG16953 standard; protein; 719 AA.

| | |
|----|--|
| AC | ABG16953; |
| XX | |
| DT | 18-FEB-2002 (first entry) |
| XX | |
| DE | Novel human diagnostic protein #16944. |

AA Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.
KW

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US008631.

PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT,

DR WPI; 2001-639362/73.

XX

aa New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.

PS Claim 20; SEQ ID NO 47312; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful for medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG03377 represent novel human diagnostic amino acid sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at http://wipo.int/pub/pdb/pdb_sequences

SQ Sequence 719 AA;

Query Match 17.7%; Score 395; DB 4; Length 719;
Best Local Similarity 30.0%; Pred. No. 2.2e-22;
Matches 121; Conservative 51; Mismatches 174; Indels 58; Gaps 18;

```
QY      3 ESDTKSISLSSSS--SRPKYICYEGCDKAYNPSSLLEOHLRTHSNDRPYKCTVD 60
Db      105 ECDKAFKPSLTHKVIHTGEKHYKC--EECGKAFTRSSSLIEHKRSHAGEKPYKC--E 160
QY      61 DCDKAFPRKSHLETHIVSHSEKPPHSCVCGKGVNSROHLKRHEITHTKSPKCTFENCQE 120
Db      161 ECGKAFSKASTLTHAKTIHAGEKPYKCECGKAFNRSSNLMHKRIHTGEKPKCECGK 220
QY      121 AFYKHGSL-RHHILSVHEKTLTKQCNKVFTRPSKLAQHKLKHGGSPAYQCD----- 172
Db      221 AFGNFSTLTGHKVIHTGEKPYKCECGKAFSWPSSLTEHKRIHAGDKP-YKCECGKTFK 279
QY      173 -----HGG-----CFKRFQTSVLQFH--IKOSHPKLCKPCKGKGVGKKG 211
Db      280 WSSLTTHKTIHTGEKPYKCECGKAFPTFSSLTGHKVIHTGEKHYKCECGKVFSSSS 339
QY      212 ISSHMLSHDSTMIKIWTCDYCDVGK-FAKKNELVEHYNI FHDGNTIPDDLKETEYKLE 270
Db      340 LTHKAIHAGE---KLYKCEEC--GKAFKMSRLSEHKRI-HTGEKP---YKCECGKAF 390
QY      271 NLDDQSKLNNLHELETEK--LKEVEDEDEEDS--LDEKRSVDVRSDSMSAQRSIKFT 325
Db      391 SWV---SVLNKHKKIHAGKKFPYKCECGKDFNQSSHLTHKRIHTGEKPYKCECGKAF- 446
QY      326 ASLEGSKSVSKLISNSGKINCPRKNCDRMFESREYDLRHLKWH 369
Db      447 -SKASTLTHAKTIHAGEKPYKC--EECGKAFNRSSNLMHKRIH 487
```

Search completed: July 22, 2004, 16:46:08
Job time : 56 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2004, 16:44:10 ; Search time 22 Seconds

(without alignments)
1801.405 Million cell updates/sec

Title: US-09-831-804-3

Perfect score: 2229

Sequence: 1 MSESDETKSSLSLSSSSSS.....PLYKKAMDLPNETSVISR 412

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1 | 837.5 | 37.6 | 429 | 2 | S20050 |
| 2 | 495 | 22.2 | 374 | 2 | T37676 |
| 3 | 400 | 17.9 | 803 | 2 | S26823 |
| 4 | 396.5 | 17.8 | 595 | 2 | G02075 |
| 5 | 395.5 | 17.7 | 363 | 2 | I38937 |
| 6 | 393.5 | 17.7 | 1191 | 2 | S35305 |
| 7 | 375.5 | 16.8 | 1350 | 2 | S00647 |
| 8 | 374.5 | 16.8 | 710 | 2 | I48668 |
| 9 | 373 | 16.7 | 542 | 2 | A54661 |
| 10 | 372 | 16.7 | 651 | 2 | B32891 |
| 11 | 372 | 16.7 | 393 | 2 | JN0533 |
| 12 | 364.5 | 16.4 | 428 | 2 | G01496 |
| 13 | 364 | 16.3 | 438 | 2 | A32891 |
| 14 | 363.5 | 16.3 | 615 | 2 | S06546 |
| 15 | 359.5 | 16.1 | 427 | 2 | A35659 |
| 16 | 354 | 15.9 | 693 | 2 | I37570 |
| 17 | 353.5 | 15.9 | 591 | 2 | S65088 |
| 18 | 351 | 15.7 | 474 | 2 | I54338 |
| 19 | 348.5 | 15.6 | 439 | 2 | S06556 |
| 20 | 348.5 | 15.6 | 707 | 2 | S68858 |
| 21 | 345.5 | 15.5 | 335 | 2 | JC1441 |
| 22 | 345.5 | 15.5 | 594 | 2 | T12488 |
| 23 | 344 | 15.4 | 337 | 2 | S60520 |
| 24 | 344 | 15.4 | 580 | 2 | A37107 |
| 25 | 343 | 15.4 | 686 | 2 | A34612 |
| 26 | 342 | 15.3 | 488 | 2 | S47072 |
| 27 | 341 | 15.3 | 336 | 2 | S06578 |
| 28 | 341 | 15.3 | 347 | 2 | S00549 |
| 29 | 340 | 15.3 | 339 | 2 | JC1442 |

| | | | | | | |
|----|-------|------|-----|---|--------|--------------------|
| 30 | 339 | 15.2 | 411 | 2 | S10245 | finger protein, te |
| 31 | 338 | 15.2 | 247 | 2 | S47070 | finger protein HZF |
| 32 | 338 | 15.2 | 383 | 2 | C32891 | finger protein 9' |
| 33 | 337.5 | 15.1 | 728 | 2 | A48830 | probable transcrip |
| 34 | 337 | 15.1 | 469 | 2 | I38600 | zinc finger protei |
| 35 | 337 | 15.1 | 595 | 2 | UC7779 | Kruppel-associate |
| 36 | 336 | 15.1 | 553 | 2 | S22954 | finger protein zfp |
| 37 | 335.5 | 15.1 | 428 | 2 | S03677 | finger protein (cl |
| 38 | 334.5 | 15.0 | 280 | 2 | S06572 | finger protein (cl |
| 39 | 332 | 14.9 | 536 | 2 | S06548 | finger protein (cl |
| 40 | 332 | 14.9 | 701 | 2 | T14757 | hypothetical prote |
| 41 | 331 | 14.8 | 794 | 2 | S59069 | zfp protein - mus |
| 42 | 330.5 | 14.8 | 732 | 2 | S47073 | finger protein HZF |
| 43 | 329.5 | 14.8 | 496 | 2 | T08674 | probable finger pr |
| 44 | 329 | 14.8 | 589 | 2 | I38598 | zinc finger protei |
| 45 | 329 | 14.8 | 728 | 2 | A54603 | transcription fact |

ALIGNMENTS

RESULT 1

S20050 transcription factor IIA - yeast (Saccharomyces cerevisiae)

N/Alternate names: protein P9677.9; protein YPR186c; TFC2 protein

Query Match 37.6%; Score 837.5; DB 2; Length 429;

Best Local Similarity 43.7%; Pred. No. 1.2e-43;

Matches 174; Conservative 62; Mismatches 135; Indels 27; Gaps 11;

| | | |
|----|-----|---|
| QY | 1 | MSESDETKSSLSLIS-SSSSSRPKYICTYGCCKANRBPDLRQH-LRTSNDPYYK 57 |
| DB | 23 | ISRSSSSLSLNTSTRSSSNRPXTYFCDDGCKAFRRPSILTEHQLSVQGLRAFOC 82 |
| QY | 58 | TVDDCDKAFPRKSHLETHIVSHSEKKPPHCVSGKGVNSRQHLKHEHTHTSSFKCTFN 117 |
| DB | 83 | --DKAKSFYKSHIERHLTYHSDTKPQCSCGCVTTROOLKHEVHTHSFICPERG 140 |
| QY | 118 | COEAFYKQSLRHLHLSVHEKTLTKCCCNKVFTRPSKLAKHKG--SPAYCCDHFG 175 |

Db 141 CNLRFRKHPOLRAHIIISVHLHKLTCPCNKSFPQPYRLRNHISKHHDPVENPYQCTPAG 200
QY 176 CFNFQOTWSVLQPHIKOSHFKLCPKCGKGVCGKGLSHMLSHDSTMIKWTCDYCDV 235
Db 201 CCKEPRFMSQLOQHIKNDHFKLCPICSKPCVGENGLQMMIITHDLSLVTKMKCHICPD 260
QY 236 GKRAKKELVENHNIIF-DSNIIPDL-LKETEVKYLENLLDQSKLNHLELETKLVE 293
Db 261 MFSFRKIDLLTHYGSITHBEDIPLEIKYKLSDIQOL--VQDHVQLGN-----SKHSNE 312
QY 294 EDEDEEDSDLSDEKSDVRSMSAQRISIKSFTASLE-GSKSVSKLISNSGKLTNCPKNC 352
Db 313 QDEBKISNRKRRKRLTENNVEFLQNHVPLEKRLSEGENGLNLLNTVGRKXRCFPNNC 372
QY 353 DRMFSERYDLRRHL--KMHDDNLQRIEFLNISTEKEE 387
Db 373 SRTEFKTEKEKHIIDKAKVHELKLIKILQ-----EKKEE 404

RESULT 2

T37676
zinc-finger protein - fission yeast (Schizosaccharomyces pombe)
C/Species: Schizosaccharomyces pombe
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 01-Dec-2000
C/Accession: T37676
R/McDougall, R.C.; Rajadream, M.A.; Barrell, B.G.; Davis, P.; Churcher, C.M.
submitted to the EMBL Data Library, October 1999
A/Reference number: Z21736
A/Accession: T37676
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-374 <MCD>
A/Cross-references: EMBL:AL132675; PIDN:CA59689.1; GSPDB:GN00066; SPDB:SPAC144.09C
C/Genetics:
A:Gene: SPAC144.09C
A:Map position: 1
A:Insertions: 50/3
C/Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

Query Match 22.2%; Score 495; DB 2; Length 374;
Best Local Similarity 32.8%; Pred. No. 5.7e-23;
Matches 123; Conservative 58; Mismatches 160; Indels 34; Gaps 13;
QY 15 SSSSSSRPKYITCTYEGCDKAYNRPDLBQHLRTHSNDRPYKCTVDDCKAFPRKSHLET 74
Db 13 SKNLRSAAKIPHCPEYEECGKYSRPSLLEQHLRTHSNRPVCDYTGCSKAFYKXSHLKI 72
QY 75 HIVHSEKKPRHGSV--CGKGVNSRQHLKREHITN--TKSPKCTPENQGEAFYHQSLRH 130
Db 73 HKRCHTNVPRPSCHYDGDCAQFYTOQHLERHIEVRRKPYACTWEGCDCEGSKHQULRS 132
QY 131 HILSVHEKTL-----TCQCNKVFTRPSKLAQHLKHHGSPAYOCDHPGC--FKNFQOTWS 184
Db 133 HISACHTHLLYPTCYDCELRPAFKQLQNHVNVAHKKIISYSPHESCVGHGEFEKMS 192
QY 185 VLQPHIKOSHFKLCPKCGKGVCGKGLSHMLSHDSTMIKWTCDYCDV---GKPAK 240
Db 193 QLOHNIIRAHVP--GCSICGRQFKTAHLRHHVVLH--QTLLEERKTYCEPMEGCKKSPTR 249
QY 241 KNEIIVENHNIIFDGNIP---DDLKETEYK--KLEMLLDQSGSKLNHLELETKLVEDE 296
Db 250 SSALKKRIHLSVIEGMAWPHCDSCGTFGKIMLQHLERGT--CKAAHPYINEGCIRKDG 308
QY 297 EDEEDSDLSDEKSDVRSMS--AQRISIKSFTASLEGSKSVSKLISNSGKLTNCPKNC 354
Db 309 IEGVALIHQKEKEELSSNLVSDAKKILNEVVG--HGK-----EAREVSCSPPECNV 358
QY 355 MFSREYDLRRHLKWH 369
Db 359 RFKRLYDMHRHINSH 373

RESULT 3

S26823
zinc finger protein ZNF43 - human
N/Alternate names: zinc finger protein kox27
C/Species: Homo sapiens (man)
C/Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 01-Dec-2000
C/Accession: S26823; 137967; S10416
R/Lovering, R.; Trowsdale, J.
Nucleic Acids Res. 19, 2921-2928, 1991
A/Title: A gene encoding 22 highly related zinc fingers is expressed in lymphoid cell 13
A/Reference number: S26823; X01D:91279444; PMID:1711675
A/Accession: S26823
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-803 <LOV>
A/Cross-references: EMBL:X59244; NID:G38031; PIDN:CAA11932.1; PID:G38032
R/Thiesen, H.J.
New Biol. 2, 363-374, 1990
A/Title: Multiple genes encoding zinc finger domains are expressed in human T cells.
A/Reference number: 137949; WUID:9145339; PMID:2288909
A/Accession: 137967
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 476-531 <THL>
A/Cross-references: EMBL:X52358; NID:G34160; PIDN:CAA36584.1; PID:G930090
C/Genetics:
A:Gene: GDB:ZNF43; HMR6
A/Cross-references: GDB:128653
A/Map position: 19p13.1-19p12
C/Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C/keywords: DNA binding; zinc finger

Query Match 17.9%; Score 400; DB 2; Length 803;
Best Local Similarity 31.9%; Pred. No. 7.5e-17;
Matches 121; Conservative 53; Mismatches 171; Indels 34; Gaps 19;
QY 2 SESDETGISISLSSSS--SRPKYITCTYEGCDKAYNRPDLBQHLRTHSNDRPYKCTV 59
Db 367 TEGCAFRRSNNLTJHKKIHTEKKPYKC--EEGKAFKWSKLTJHKKIHTGEKPYKC-- 422
QY 60 DDCDKAFPRKSHLETTHIVSHSEKKPRHGSVCGKGVNSRQHLKREHITN--KSPKCTPEN 117
Db 423 EECGKAFNWSSTLTJHKKIHTEKPYKCEYCGKAFNFSNLTJHKKIHTTAKPYKC--EE 480
QY 118 QGEAFYHQSL--RHIIISVHEKTLTCQCNKVFTRPSKLAQHLKHHGSPAYOCDHPGC 176
Db 481 CGKAFSSSNLTJHKKIHTEKPYKCEGKAFKWSKLTJHKKIHTTGERK--YKCEB--C 537
QY 177 FKNFQOTWSVLQFH--IKOSHFKLCPKCGKGVCGKGLSHMLSHDSTMIKWTCDYCD 234
Db 538 GKANHHSILTKHKKIHTEKPYKCEBGKAFQSSNLTJHKKIH--TGKFFYKCEBG-- 593
QY 235 VGR-FAKKNELVEHYNIFHDGNIPDDLKETEYKLEMLLDQSGSKLNHLELETKLVE 293
Db 594 -GKAFQSSNLTJHKKI-HTGKRP--YKCEGKAFNFSNLTJHKKIH--TGKFFYKCEBG-- 647
QY 294 E-DEDEEDSDLSDEKSDVRSMS--MSAQRISIKSFTASLEGSKSVSKLISNSGKLTNCPK 350
Db 648 ECGKAFKWSSTLTJHKKIHTEKPYKCEBGKAF--KLSSTLTJHKKIHTEKPYKCKK-- 704
QY 351 NCDRMFSREYDLRRHLKWH 369
Db 705 -CGKAFNRPNSLTJHKKIH 722

RESULT 4

G02075
transcription repressor zinc finger protein 85 - human
C/Species: Homo sapiens (man)
C/Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 01-Dec-2000
C/Accession: G02075
R/Poncellet, D.A.
submitted to the EMBL Data Library, September 1995

A:Reference number: G09169
A:Accession: G02075
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-595 <PON>
A:Cross-references: EMBL:U53576; NID:g1017721; PIDN:AAA79179.1; PID:g1017722
C:Genetics:
A:Gene: GDB:ZNF85
A:Cross-references: GDB:132279
A:Map position: 19p12-19p12
C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

Query Match 17.8%; Score 396.5; DB 2; Length 595;
Best Local Similarity 31.1%; Pred. No. 8.8e-17;
Matches 120; Conservative 48; Mismatches 160; Indels 51; Gaps 17;

QY 10 ISSLSSSSS-SRPKKYICTYEGCDKAYNRPISLEQHLRPHNSNRPKCTVDDCDKAFR 68
D 186 ISCTEHSRIHTRVNFYKC--EECGKAFNWSSTLTGKRIHTEGKPYKC--EECGKAFNQ 241
QY 69 KSHLETHIVSHSEKPPHCVCVGKGVNSROHLKKEHITHT--KSFCTFENCCEAFYKHQ 126
D 242 SSNLIKHKKIIHTGKPKCECGKAFNRFSTLTGKIIHTEGKPYKC--KECGKAFNRSS 299
QY 127 SL-RHHTLSVHEKLTGKCNKVTRPSKLAQKTLKHGSPAYQCDHPGCFKNFQTSV 185
D 300 TLTHRKIIHTEGKPYKCECGKAFKQSSNLTGKIHTGKPYKCK--CGKAFNQSAAH 356
QY 186 LQFFH-IKOSHPRKLCPCCKGCVGKGLSSHMLSHDSTPMIKITWTDYCDYGFAPKNE 243
D 357 LTHHEVHTGEEKPYKCKCGKAFNHFSLTTHKIIH--TGEKPYKCKEC--GKAFNQS 411
QY 244 LVEHYNIFHDGNIPTDLIKETEVKLEMLDQSKLNHLELTELKVEDEDEDESL 303
D 412 TLTHKIIHTEG-----KPYKSECEKAFNQSGL-----TEHKKIHTGKPYE--- 455
QY 304 DEKRSVDPRSDMSMAQSRISKFTASL-----EGSK-----SVSKLISNGKINCPKN 350
D 456 CEKCGKAFNQSSTLTGKRIHTEGKPYKCECGKGFPMSTLTGKIIHTEGKPYKC--E 513
QY 351 NCDRMFSREYDLRRLKWH 369
D 514 ECGKAFNQSSTLTGKRIH 532

RESULT 5
138937
DNA/RNA-binding protein - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 21-Jul-2000
C:Accession: I38937
R:Drew, P.D.; Nagle, J.W.; Camling, R.D.; Ozato, K.; Biddison, W.E.; Becker, K.G.
Gene 159, 215-218, 1995
A:Title: Cloning and expression analysis of a human cDNA homologous to Xenopus TFIITIA.
A:Reference number: I38937; MIMD:95347600; PMID:7622052
A:Accession: I38937
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-363 <RES>
A:Cross-references: EMBL:U02072; NID:g644870; PIDN:AAA75623.1; PID:g644871
C:Superfamily: transcription factor TIIA

Query Match 17.7%; Score 395.5; DB 2; Length 363;
Best Local Similarity 31.1%; Pred. No. 5.9e-17;
Matches 99; Conservative 54; Mismatches 96; Indels 69; Gaps 16;

QY 8 KSISST-----ISSSSSR-----PKYICTEYEGCDKAYNRSLEQHLRTSNDP 53
D 7 ESSVSLTIDAFIAAGESSAPTPRPALPRFICSPDSCANSYKAMKIDAHLCRTGGER 66
QY 54 PYKCTVDDCDKAFRRKSHLETHIVSHSEKPPHCVCV--CGKGVNSROHLKRIH-ETHT--- 107
D 67 PFVCDYEGCGKAFIRIYHLSRHILTHTEGKPYKCANCGDCKNTKSNLKKFERKHEHQ 126

QY 108 TKSFKCTFENCCEAFYKHQSLR--HHILSVHEKLTGCKO--CNKVETPSKLAQKIKHHG 164
D 127 QKQYICGFEDOCKTFKKHQQLKHQCCHTNEPLFLFKTQDEGCGKHFASPSKILAKAKAHEG 186
QY 165 GSPAYQCDHPGCFKNFQTSVLFQHIKQSHPK----- 196
D 187 -----YVC--QKGCSPVAKTWTLELKHVRETHKEELICVCRKTFKRDYIKQMKHTAPER 241
QY 197 --LKCPK--CGKGVGKGLSSHLS--HDSSTMIKITWTDYCDYK--PAKKNELVEHYN 250
D 242 DVCCPREGGGRITVTYFNLIQSHLSPHESS---RPVCEHAGCGKTFAMKQSLTRH--AV 297
QY 251 FHDGNIPTDLIKETEYVK 268
D 298 VHD---PDKKMKLKYK 312

RESULT 6
535305
zinc finger protein ZNF91 - human
C:Species: Homo sapiens (man)
C>Date: 31-Dec-1993 #sequence_revision 02-Jun-1994 #text_change 05-Nov-1999
C:Accession: S35305
R:Beilströmd, E.V.; Marine, J.C.; Ried, T.; Lecocq, P.J.; Riviere, M.; Amemiya, C.; Po
EMBO J. 12, 1363-1374, 1993
A:Title: Clustered organization of homologous KRAB zinc finger genes with enhanced exp
A:Reference number: S35305; MIMD:93223677; PMID:8467795
A:Accession: S35305
A:Molecule type: mRNA
A:Residues: 1-1191 <BEL>
A:Cross-references: EMBL:L11672; NID:g186773; PIDN:AAA59469.1; PID:g186774
A:Note: the authors translated the codon GCA for residue 750 as Thr and GCT for residu
C:Genetics:
A:Gene: GDB:ZNF91; HPF7; HRF10
A:Cross-references: GDB:132284
A:Map position: 19p12-19p12
C:Keywords: DNA binding; zinc finger

Query Match 17.7%; Score 393.5; DB 2; Length 1191;
Best Local Similarity 35.8%; Pred. No. 2.8e-16;
Matches 98; Conservative 40; Mismatches 109; Indels 27; Gaps 12;

QY 20 SRPKYICTYEGCDKAYNRPISLEQHLRPHNSNRPKCTVDDCDKAFRRKSHLETHIVSH 79
D 905 TREKPYKC--EECGKAFSQSPSHLTGKRIHTEGKPYKC--EECGKAFSQSSTLTGKRIH 960
QY 80 SEKKPPHCVCVGKGVNSROHLKKEHITHT--KSFCTFENCCEAFYKHQSL--RHHTLSVA 136
D 961 TGEKPYKCECGKAFNRSSSTLTGKRIHTEGKPYKC--EECGKAFSQSSTLTGKRIH 1018
QY 137 EKLTLGKCNKVTRPSKLAQKTLKHGSPAYQCDHPGCFKNFQTSVLFQHIKQSHPK 136
D 1019 EKPYKCECGKAFNRSSSTLTGKRIHTEGKPYKCKE--CGKAFISSSTLNGH--KRILTR 1074
QY 197 ---LKCPKCGKGVGKGLSSHLS--HDSSTMIKITWTDYCDYGFAPKKNELVEHYNIFHD 253
D 1075 EKPYKCECGKAFNQSSTLTGKRIH--TGEKPYKCGEC--GKAFKSSALTKKIIHT 1129
QY 254 GNIPDDLKETEYVKLEMLDQSKLNHLELET 287
D 1130 GE-----KPYKCEKCKAFNQSSTLTGKRIH 1157

RESULT 7
S00647
finger protein - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C>Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 31-Dec-1993
C:Accession: S00647
R:Pruiz i Altaba, A.; Perry-O'Keefe, H.; Melton, D.A.
EMBO J. 6, 3065-3070, 1987
A:Title: Xfin: an embryonic gene encoding a multifingered protein in Xenopus.

A;Reference number: S00647; MUID:88082679; PMID:2826129
A;Accession: S00647
A;Molecule type: mRNA
A;Residues: 1-1350 <RUI>
A;Cross-references: EMBL:X06021
A;Note: it is uncertain whether Met-1, Met-11, Met-16, Met-38, or Met-39 is the initiator
C;Genetics:
A;Gene: fln
C;Keywords: DNA binding; zinc finger

Query Match 16.8%; Score 375.5; DB 2; Length 1350;
Best Local Similarity 27.3%; Pred. No. 4e-15;
Matches 119; Conservative 65; Mismatches 171; Indels 81; Gaps 20;

QY 11 SLSISSSSSPRPKYICTYEGCDAYNRPSSLLEOHLRTHSNDPRYKCTVDDCDYAFPRKS 70
DB 340 SFLNQQTSHREKPYLCSH--CNKGFIQNSDLVGHFRHTGTERPYQCA--ECHGFIQKS 395
QY 71 HLETHIVSHSKRPKPHSCGVNRSQHLKRHEITHT--KSFKCTFENCOEAFYKQSL 128
DB 396 DLVHHLRTHTEKPKFCHCDKFTERSALAKHORHTGTERPYKCS--DCKEFTQSNL 453
QY 129 -RHHLISVHEKTLTCKOCNKVFTRP SKLAQHLKH-----HGSPAYOCDHPGCFKN 179
DB 454 ILHQRIHTGERPYKCTLCDRFIQNSDLVKHQVHANPLSDPHANSPHKCSK--CDLT 511
QY 180 FQTVSVLQFHIK--QSHPKLCPKCGKCGVCGKGLSHMLSH-----DD 221
DB 512 FSHWSTFWKSHKSHGEEKFQCAECKGFTQKSDLVKILIRVHTGTERPKCLLCKKSPSON 571
QY 222 STMKIM-----TCDCYDVGKPAKKNELVHEVYNIFHDGNTPDLLKETEYKLENL 272
DB 572 SDLHKHMRHTGTERKPPCYTCD--KSTERSALIKHNR--HTGERPH-----KCSVCKG 623
QY 273 LDQSKLNINLHELETEKLKVEDEDEDSIDEKRSVDYRSDMSAQR-----S 320
DB 624 FIQGSAL-----TKHSRHTGTERKPYCTQCGKSFQNSDLVKHQRIHTGTERPYHCTEC 676
QY 321 IKSTFASLEGSKSVSKLISNGKK--INCPKNNCDMFREYDLRHLKMH--DNLQRIES 378
DB 677 NKRFTH---EGSLVYHNRTHSGEKYRCPQ--CEKTFIQSSDLVHNLVHNGENPPATA 731
QY 379 FLN-SIEKEETPEGER 393
DB 732 PHEILIRENLTREBP 747

RESULT 8
148668
zinc finger protein 51 - mouse
N;Alternate names: finger protein zfc12
C;Species: Mus musculus (house mouse)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 08-Dec-2000
C;Accession: I48668; A40984
R;Burke, P.S.; Don, J.; Wolgemuth, D.J.
Mamm. Genome 5 387-389, 1994
A;Title: Zfp-51, a murine zinc finger encoding gene mapping to the t-complex region of C
A;Reference number: I48668; MUID:94319090; PMID:8043957
A;Accession: 148668
A;Status: preliminary;
A;Molecule type: mRNA
A;Residues: 1-710 <RES>
A;Cross-references: EMBL:X74855; NID:9488832; PIND:CAA52847.1; PID:9488833
R;Crossley, P.H.; Little, P.F.R.
Proc. Natl. Acad. Sci. U.S.A. 88, 7923-7927, 1991
A;Title: A cluster of related zinc finger protein genes is deleted in the mouse embryonal
A;Reference number: A40984; MUID:91376058; PMID:1680234
A;Accession: A40984
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 672-710 <CRO>
A;Cross-references: GB:M74235
C;Genetics:

A;Gene: Zfp-51
C;Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C;Keywords: DNA binding; zinc finger

Query Match 16.8%; Score 374.5; DB 2; Length 710;
Best Local Similarity 29.6%; Pred. No. 2.3e-15;
Matches 104; Conservative 54; Mismatches 140; Indels 53; Gaps 15;

QY 25 YICTYEGCDAYNRPSSLLEOHLRTHSNDPRYKCTVDDCDYAFPRKSHLETHIVSHSKRP 84
DB 311 YEC--KECGSFIELSLKHKHYRHTGTERPKCI--CDSFTTTLTKTHQKHTGTERK 366
QY 85 FHCSCGVNRSQHLKRHEITHT--KSFKCTFENCOEAFYKQSLRHHLS--VHEKTLT 141
DB 367 YKRECKSPKSHSHLRHQNVTGERPYK--KECKSPHESATLAEHESHTGTERK 424
QY 142 CKQCNKVFTRPSPKLAQHLKHNGSPAYOCDHPGCFPNFQTVSVLQFH--IKQSHPKLC 199
DB 425 CRECDKSFQRAVLRHNRVHTGERPYCKE--CGKSFYTCSTLRHQTHQTKKPYKC 482
QY 200 PKCGKCGVCKGLSHMLSHDSTMIKMTCDYDVGKPAKKNELVHEVYNIFHDGNTPD 259
DB 483 IECKSPFNNSYLTTHQVHSGE--KPYRCKED--KSPFSCSTLKAHQSL--HTGTERK 537
QY 260 LKETEYKLENLDDQSKLNINLHELETEKLKVEDEDEDSIDEKRSVDYRSDMSAQR 319
DB 538 CI-----ECDKSTFYVSHLRTHQVHTGERPFCED-----KSFIRSHLRHQ 583
QY 320 SIKSTFASLEGSKSVSKLISNGKK--INCPKNNCDMFREYDLRHLKMH 369
DB 584 KI-----HSGEKPYKC--RDCDISPSQISLNRHQA 614

RESULT 9
A54661
zinc finger protein ZNF41 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 02-Jun-1995 #sequence_revision 02-Jun-1995 #text_change 01-Dec-2000
C;Accession: A54661; I54227
R;Franze, A.; Archidiacono, N.; Rocchi, M.; Marino, M.; Grimaldi, G.
Genomics 9, 728-736, 1991
A;Title: Isolation and expression analysis of a human zinc finger gene (ZNF41) located
A;Reference number: A54661; MUID:91244317; PMID:2037297
A;Accession: A54661
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-542 <FRA>
A;Cross-references: GB:M6548; GB:M92443; NID:9340443; PIND:AAA61312.1; PID:9340444
C;Genetics:
A;Gene: GDB:ZNF41
A;Cross-references: GDB:125865; OMIM:314995
A;Map position: Xp11.3-Xp11.23
C;Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

Query Match 16.7%; Score 373; DB 2; Length 542;
Best Local Similarity 29.0%; Pred. No. 2.1e-15;
Matches 115; Conservative 59; Mismatches 156; Indels 66; Gaps 21;

QY 14 ISSSSSPRPKYICTYEGCDAYNRPSSLLEOHLRTHSNDPRYKCTVDDCDYAFPRKSHLE 73
DB 163 IHOKHTGTERKPYEC--NECGAFPRKSLRMRQRIHTGTERPYVCA--DCGAFYQKSHFN 218
QY 74 THIVSHSEKRPKPHSCGVNRSQHLKRHEITHT--KSFKCTFENCOEAFYKQSLRHH 131
DB 219 THQRIHTGTERPYECSDCKSTKKSQLVHQRHTGTERPYCT--ECKGVFTRHN----- 272
QY 132 ILVVEKTLT-----CKQCNKVFTRPSPKLAQHLKHNGSPAYOCDHPGCFPNFQTVSV 185
DB 273 -LTHQKHTGTERKPYMAECGAFYDQSNLILKHQTHGTERK--YCN--GCGKAFIWKSR 328
QY 186 LQFHQKSH---PKLCPKCGKCGVCKGLSHMLSHDSTMIKMTCDYDVGK--FAKK 241
DB 329 LKIH-QKSHIGERHVECKDCGAFYQKSTLSVHORIH--TEKPYVCEC--GGAFTQK 382

QY 242 NEVEHNIIFHDGNIIPD-----LKTEVYKLENIILDOGSKLNNIHF----- 284
Db 383 SHFIAHRT-HTGEKPYECSDCGKCFTKKSQ-LRVHQKHTEGKPNICAECKAFTRSN 440
QY 285 -LETEKLVEDEDEDS-----LDEKSDVRSMSAQRI-----KSTALBESKS 333
Db 441 LITQKHTHREKPEECBDCGCTTFTWKSRLNTHQKSHGERHYECSKCAFIQ--KATLS 498
QY 334 VSKLISNGKKNPCNNCDPMFSGREYDLRHLKMH 369
Db 499 MHQIHTGKKPYAC--TECKAFIDRSNLIKHMGM 532

RESULT 10

B32891
finger protein 2, placental - human
C/Species: Homo sapiens (man)
C/Date: 20-Oct-1989 #sequence_revision 20-Oct-1989 #text_change 01-Dec-2000
C/Accession: B32891
R/Bellefroid, E.V.; Lecocq, P.J.; Benhida, A.; Poncelet, D.A.; Belayew, A.; Martial, J.A.
DNA 8, 377-387, 1989
A/Title: The human genome contains hundreds of genes coding for finger proteins of the
A/Reference number: A32891; MUID:89377476; PMID:2505992
A/Accession: B32891
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-651 <BEL>
A/Cross-references: GB:M27878
C/Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C/Keywords: tandem repeat

Query Match 16.7%; Score 373; DB 2; Length 651;
Best Local Similarity 27.4%; Pred. No. 2,6e-15;

Matches 128; Conservative 57; Mismatches 146; Indels 136; Gaps 24;

QY 23 KKVICYEGCDKXVNPISLEQHLRTHSNRPYKCTVDDCDKAFRRKSHLETHVSH--- 79
Db 174 KPYNCQ--CGKAFSQSQTLSHQRTHTGKPYEC--GECGKAFRRKSHLSHMRHTG 229
QY 80 -----SEK-----KPHCVGCGKGVNSRQHLKRHEITH--TKSPK 112
Db 230 KPYNCGCGRAFESEKSNLTHQRTHTGKPYECGCGKAFRRKSHLVTHRHTGTPRG 289
QY 113 CFPEHCQAFY-KHQSILRHILSVHEKTLCKQCNVFTPSKLAQHKLKHGSGP----- 167
Db 290 CS--DCKAFPEKSEHLRHQTHHTGKPYECSCRAFRSRSLNHQRTHTGKPHGCI 347
QY 168 -----AYQCDHG-----CFKNPOTWSVLQPHIKOSHPRK--CPKCG 203
Db 348 QCGKAFSQSKSHLSHQRTHTGKPFICSKCGKAFRRKSHLVTH--QRTHTGKPYECSECG 406
QY 204 KGCVGKGLSHMLSHDSTMIRKTCYCDVGVK-PAKKNELVEHNIIFHDGNIIPDIL- 261
Db 407 KAFSEKLSLTNRHRIH---TGKPYECSC--GKAFQCSHLSHRT-HTGEKPYECSE 460
QY 262 -----KETEYKLENIILDOGSKLNNIHELETEKLVEDEDEDS 300
Db 461 CGKAFGEKSLATHQRTHTGKPYECRDEKAFSQSQNT-HQ-----RLHTGKPYE 513
QY 301 DSLDERSDVRSMSAQRSIKSFASLSEGSVSKLISNGK-INCPKNNCDPMFSGE 359
Db 514 GSGCRK-----AFPEKSELIRLIRHTGKPYEC--NECKRAFRK 552
QY 360 YDIRHLLKWHNDLORISFPLNSIEKEPTPEGEPLVKKARMLDLPNE 406
Db 553 SSLINH-----QRIHTGKPEECSE--GKAFRRKSH--LIRHQ 587

RESULT 11

JN0533
finger protein pMLZ-4 - mouse
C/Species: Mus musculus (house mouse)

C/Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 01-Dec-2000
C/Accession: JN0533
R/Brady, J.P.; Piatigorsky, J.
Gene 124, 207-214, 1993
A/Title: Cloning and characterization of a novel zinc-finger protein encoding cDNA from
A/Reference number: JN0533; MUID:93185925; PMID:8444344
A/Accession: JN0533
A/Molecule type: mRNA
A/Residues: 1-393

A/Cross-references: GB:M98502; NID:g200406; PIDN:AAA39949.1; PID:g200407
A/Experimental source: eye lens
C/Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C/Keywords: DNA binding; eye lens; zinc finger
F/1-62/Region: acidic
F/63-391/Region: zinc finger

Query Match 16.7%; Score 372; DB 2; Length 393;
Best Local Similarity 29.4%; Pred. No. 1,7e-15;
Matches 110; Conservative 57; Mismatches 163; Indels 44; Gaps 17;

QY 13 LISSSSSRKKYICTEYEGCDKXVNPISLEQHLRTHSNRPYKCTVDDCDKAFRRKSHL 72
Db 49 LVHKEAHAGIRYHICQ--CGKAFSQSDINRHQKHTGDRPYKCY--ECGKGFSSSHL 104
QY 73 ETHIVSHSEKPPHCVGCGKGVNSRQHLKRHEITH--KSPKCFPEHCQAFYKQSL-R 129
Db 105 IQHQRTHTGERPDCNCEGKSPFRSSHLTHQRTHTGKPHKCT--ECAAASAPHLIQ 162
QY 130 HHLISVHEKTLTKQCNVFTPSKLAQHKLKHGSPAYQCDHPGCFKNPOTWS--VLQ 187
Db 163 HQRTSGEKPYECECEGKSFSSHLAQHRTHTGERP-YECHE--CGRFSERSDLIKH 219
QY 188 FHTKQSHPRKLCGKCGKGVNSRQHLKRHEITH--KSPKCFPEHCQAFYKQSL-R 247
Db 220 YRVHTEGKPYECGKSNFSQNSDLVRHRAH---TGKPYHNECG-ENFSRISHLVQH 275
QY 248 YNIFHDGNIIPDILKETEYKLENIILDOGSKLNNIHELETEKLVEDEDS-----EDEDS 302
Db 276 QRT-HTGE-----KYECTACGKSSRSRSHLTHQKHTHTGKPYECNEMRSGERSDL 328
QY 303 LDEKSDVRSMSAQRSIKSFASLSEGSVSKLIS-----NSGKR-INCPKNNCDPMF 357
Db 329 IKHQRTHTGKPYECVQCGKGFQSQ-----SNLTHQRTHTGKPYEC--TECDKSF 379
QY 358 REYDLRHLKWHDD 371
Db 380 RSSALIKHRVHTD 393

RESULT 12

G01496
transcription factor IIA - human (fragment)
N/Alternate names: TFIITIA
C/Species: Homo sapiens (man)
C/Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 05-Nov-1999
C/Accession: G01496
R/Becker, K.G.
submitted to the EMBL Data Library, August 1994
A/Reference number: G07520
A/Accession: G01496
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-338 <BEC>
A/Cross-references: EMBL:U14134; NID:g551534; PIDN:AAA21873.1; PID:g551535
A/Genes: GDB:GTF3A; TFIITIA
A/Cross-references: GDB:434744; OMIM:600860
A/Map position: 13q12.3-13q13.1
C/Superfamily: transcription factor IIA

Query Match 16.4%; Score 364.5; DB 2; Length 338;
Best Local Similarity 29.0%; Pred. No. 4,1e-15;
Matches 89; Conservative 49; Mismatches 102; Indels 67; Gaps 11;

```

QY      8 KSIISL-----ISSSSSR-----PKYICTCYGCDKAYNRBSLEQHLRTSNDR 53
QY      7 ESVSLSLTAAFLTAAGSSSAPTPRPALPRFLICSPFDCSANSKMKLDAHLCCKHTGR 66
QY      54 PYKTVDDCDKAFPRKSHLETHIVSHSEKRPFHCSV--CGKVNRSOHLKRH-ETH--- 107
QY      67 PFVCDYEGCGCAFRDYLNRHILTHTGKRPFCVCAANGCDQKFNPKSLKKHFERKIENG 126
QY      108 TKSKCPFPENCOEAFYHQSLR-----HHILSVH 136
QY      127 QKQITCSBEDCKKPFKQHQQLKTHQCOHTNEPLFLKCCQKCSVATWTWELLKHVETH 186
QY      137 EKITLTCKQCNKVFTRPSKLAQHKLKHGGSPAYOCDHPGCFKNFQTVSLQFIHKOSHPP 196
QY      187 KEELLCECRKTFPRKDYLIKQMKHTHAPEROVCRCPREGCGRTVTYFNLIQSHILSPHEE 246
QY      197 LK---CPK--CGKGCYGGKGLSSHMLSHD-DSTMIKITWTCDYCDYGVFAKKNELVHYNT 250
QY      247 SRPVCERHAGCGKTFAMKQSLTRHAVVHDPDKKKKL-----KVKKSRKRSLSLASH- 299
QY      251 FHDGNIP 257
QY      300 ---GYTP 303

RESULT 13
A32891
finger protein 1, placental - human
C:Species: Homo sapiens (man)
C:Date: 20-Oct-1989 #sequence_revision 20-Oct-1989 #text_change 01-Dec-2000
C:Accession: A32891
R:Bellefroid, E.J.; Iecocq, P.J.; Benhida, A.; Poncelet, D.A.; Belayew, A.; Martial, J.A.
DNA 8, 377-387, 1989
A>Title: The human genome contains hundreds of genes coding for finger proteins of the K
A:Reference number: A32891; MUID:89377476; PMID:2505992
A:Accession: A32891
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-428 <BEL>
A:Cross-references: GB:M27877
C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C:Keywords: tandem repeat

Query Match          16.3%; Score 364; DB 2; Length 428;
Best Local Similarity 29.1%; Pred. No. 5,7e-15;
Matches 116; Conservative 50; Mismatches 155; Indels 78; Gaps 18;

QY      32 CDKAYNRBSLEBQHLRTSNDRPYKCTVDDCDKAFPRKSHLETHIVSHSEKRPFHCSVCG 91
QY      38 CGKIPNKKSNLASHQRIHTGKRPYKC--NECGKVFHNMSHLAQHRITHTGKRPYKCNCG 95
QY      92 KGVNRSOHLKRHEIHT--KSPKCTFENCOEAFYHQSL-RHHILSVHEKITLTCKQCNKV 148
QY      96 KVFNOISHLAQHRIHTGKRPYKC--NECGKVFNOISHLAQHRTHTGKRPYECNCKGKV 153
QY      149 FTRPSKLAQHKLKHGGSPAYOCDHPGCFKNFQTVSLQFI--IKOSHPKLKCPKCGKGC 206
QY      154 FSRNRYIVQHLIHTGKRP--YRCN--VCGKVFHNMISHLAQHRIHTGKRPYKCNCGKVF 210
QY      207 VGKKGSLSHMLSHDSTIMIKITWTCDYCDYGVK-FAKKNELVHYNTFHDGNIP----- 257
QY      211 SHKSSLVNHWRIH--TGKRPYKCNCG--GKVFSHKSSLVNHWRI--HTGKRPYKCNCGKC 264
QY      258 -----DDL-----KETEVKKLEMLIDGSGKINMLHELETEKLKVEDEDEDEDL 303
QY      265 VFSRNSYLAQHILTAAGEKPYKCECDKAFSGNSILVQHRRIHTGKRPYKCDGCGKVFSG 324
QY      304 D-----EKRSQVR-----SDSMSAQRST-----KSTFASLEG 330
QY      325 NSYLAVHMRITHTGEKAYKCNCGKVFGLNSSLAHHRKIHTGKRPKCNCGKAF--SMRS 382
QY      331 SKVSQKLTLSNGKTKNCPKNCDCRMFSPREYDLRRHLKMH 369

```

```

Db      : 363 SLTNHAIHTGKHKRC--NECGKLFRDMSLYVRHQRFH 419
                                                                    :|::||:
RESULT 14
506546
finger protein (clone XlCOF7.1) - African clawed frog (Fragment)
C:Species: Xenopus laevis (African clawed frog)
C:Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 01-Dec-2000
A:Accession: S06546
R:MotifField, W.; El-Baradi, T.; Mentzel, H.; Pieler, T.; Koester, M.; Poetting, A.; Knoechel, J.Mol. Biol. 208, 639-659, 1989
A:title: Second-order repeats in Xenopus laevis finger proteins.
A:Reference number: S05632; MUID:J90040698; PMID:2509712
A:Accession: S06546
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-615 <NIB>
C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C:Keywords: DNA binding; zinc finger

Query Match           16.3%; Score 363.5; DB 2; Length 615;
Best Local Similarity 27.6%; Pred. No. 9.1e-15;
Matches 119; Conservative 48; Mismatches 165; Indels 99; Gaps 18;

QY      11 SLTSSSSSRBRPKYICTEGCDKAVNRPSLRLCHLRTHSNDRPYKCTVDVDCDAFFPKS 70
        ||:::||::||::||::||::||::||::||::||::||::||::||::||::||::||
DB      128 SLVHHQRTHIGKEPFCSE--CDKCPASSSELMIHQRTHTGKEKPFSCS--ECGKCFTHHS 184
        ||::||::||::||::||::||::||::||::||::||::||::||::||::||

QY      71 HLETHIVSHSEKKPFHCISVOGKGNSRQLKRHEITHT--KSFKCTEENCCEAFYKHSL 128
        ||::||::||::||::||::||::||::||::||::||::||::||::||::||
DB      185 HFANHQMHTGKEKPFCSCKGKCFASSSDLTFFRRTRTREKTFSCS--ECGKCFGNHSL 242
        ||::||::||::||::||::||::||::||::||::||::||::||::||::||

QY      129 -----RH 130
        ||::||::||::||::||::||::||::||::||::||::||::||::||::||

DB      243 ANHMHTGEKPCCSECGKCFSSSGLTANHQRTMKNKVPFSCSACGKCFSTNSHLRH 302
        ||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY      131 HLTVHREKTLTGCKNQKVFRPSKLAQHKLHHGGSYPAYOCDHPGCFNPOTWSVLQFIH 190
        ||::||::||::||::||::||::||::||::||::||::||::||::||::||
DB      303 QMTHTGEKPFSCCECKKCRSPSNLARHQMTHTGERK--FGSSE--CGKCFPASSDLTFH 358
        ||::||::||::||::||::||::||::||::||::||::||::||::||::||

QY      191 KOSHPELK--CPKCGKGVGKK-GLSNMLSHDSTMIKIWTCDYCVGKFAKNELVE 246
        ::||::||::||::||::||::||::||::||::||::||::||::||::||
DB      359 HHTHTGEKPFSGCEGCK-CYSKSSLVHQRTHTGKEKPFSCSKDKC---FASSSEL-- 411
        ::||::||::||::||::||::||::||::||::||::||::||::||::||

QY      247 HNIPFDGINPDLLKETEVKYKLNLDOGSKYLNHLFLTEKLVKEDPRDEBDSIDX 306
        ::||::||::||::||::||::||::||::||::||::||::||::||::||
DB      412 --NIHQRTHTGEKAFCSECGKCFTHRSQLSRHOMHTGE-KPIISCPCECFVAS-- 464
        ::||::||::||::||::||::||::||::||::||::||::||::||::||

QY      307 RSIVRDSMSAQSISSFSLSGSKVS-----KLINSNGKINCPKNCQDMFSR 358
        ::||::||::||::||::||::||::||::||::||::||::||::||::||
DB      465 -SGLTHHQQAHWKPFSC-LBCGKCFSNSRNPAHQMITHTGKEKPFSC--SECKRGFSN 520
        ::||::||::||::||::||::||::||::||::||::||::||::||::||

QY      359 EYDLRRHLKWH 369
        :|||::|||
DB      521 QSILARHQMTH 531
        :|||::|||

RESULT 15
A35659
kruppel-related protein H-p1k - human
C:Species: Homo sapiens (man)
C:Date: 28-Sep-1990 #sequence_revision 18-Nov-1992 #text_change 01-Dec-2000
A:Accession: A35659
R:Kato, N.; Shimotohno, K.; VanLeeuwen, D.; Cohen, M.
Mol. Cell. Biol. 10, 4401-4405, 1990
A:title: Human proviral mRNAs down regulated in choriocarcinoma encode a zinc finger pr
A:Reference number: A35659; MUID:90318410; PMID:2115127
A:Accession: A35659
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-427 <XAT>
A:Cross-references: GB:U55422; NID:g184342; PIDN:AAA36010.1; PTD:g184343
```

C;Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

Query Match
Post: 10557

16.1%; Score 359.5; DB 2; Length 427;
32.0%; Pred No 1 1a-14.

Best Local Similarity 32.0%; Pred. No. 1.1e-14;
Matches 99; Conservative 36; Mismatches 121

Matches 99; Conservative 36; Mismatches 121; Indels 53; Gaps 13;

| | | | |
|----|-----|--|-----|
| Qy | 20 | SRPKYUJCTYEGCDKXVANNRSLLEQJLRTHSNDPRPKTVDDCDKAFRRKSHLETHVSH | 79 |
| Dp | 132 | TRVAFYFC--BAVGRFANWSSTLNNKRRJHTEKEPKYC--KECGAFNFQJSHLJRRHRIRH | 187 |
| Qy | 80 | SEKKPRHCVSQGVANSROHLKRNHEITHTKS--FKCTEENCQEAIFYKQJSL-RAHHSVH | 136 |
| Dp | 188 | TEEKPRYKCEBGCAFNQSSJTLTHNIJHTGEIPLYOC--EKCVRAFNQSKJLTHEKJLHTG | 245 |
| Qy | 137 | EKJTLTCQCKNVATPRSLAOKTKY-----HGSSPAY | 169 |
| Dp | 246 | EKYECHECGKAPNRSSTKJLTHEKYJHTGEKJLYKCECGKAFNQSJTLTHNKRJHSGKPRY | 305 |
| Qy | 170 | QCDHPGCFKXNFQJWTVSLQFH--IKOSHPRJLCKPRKCGCVGKKJGSHSLSHDSJTWIKJ | 227 |
| Dp | 306 | KCEB--CGAFKQFQSNJLTHKKIHJGEKPRYKCEBGCAFNQJSLNLTNNKXVH--JCEKP | 360 |
| Qy | 228 | WTEDYCDVCGFKAKNLEIVHVNIFPDGNIPBDJLKEJPEYKJLEJMLLOQSKJLNNHLEJLT | 287 |
| Dp | 361 | YKQGCBC--GKAPNQSALANTHKJHJTGENPHKCRBSGKVFHL-----SSKJLSTOCKJHT | 412 |
| Qy | 288 | -EKL-KVES | 294 |
| Dp | 413 | GEKJLYKCEE | 421 |

Search completed: July 22, 2004, 16:47:55
Job time : 23 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2004, 16:42:55 ; Search time 13 Seconds

(without alignments)
1650.223 Million cell updates/sec

Title: US-09-831-804-3

Sequence: 1 MESSDETKSISSLISSSSSS.....PLYKARMDLIPNETYSISR 412

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Match | Length | DB ID | Description |
|------------|-------|---------|--------|-------|-------------|
| 1 | 837.5 | 37.6 | 429 | 1 | TF3A_YEAST |
| 2 | 495 | 22.2 | 374 | 1 | TF3A_SCHPO |
| 3 | 400 | 17.9 | 803 | 1 | ZN4_HUMAN |
| 4 | 398 | 17.9 | 574 | 1 | TF3A_HUMAN |
| 5 | 397.5 | 17.8 | 423 | 1 | TF3A_HUMAN |
| 6 | 396.5 | 17.8 | 595 | 1 | ZN85_HUMAN |
| 7 | 395 | 17.7 | 616 | 1 | ZN93_HUMAN |
| 8 | 393.5 | 17.7 | 1191 | 1 | ZN91_HUMAN |
| 9 | 387.5 | 17.4 | 839 | 1 | ZN47_HUMAN |
| 10 | 385 | 17.3 | 576 | 1 | ZN43_HUMAN |
| 11 | 383 | 17.2 | 913 | 1 | ZN28_HUMAN |
| 12 | 376 | 16.9 | 570 | 1 | ZN40_HUMAN |
| 13 | 376 | 16.9 | 751 | 1 | ZN84_HUMAN |
| 14 | 375.5 | 16.8 | 1350 | 1 | XFIN_XENLA |
| 15 | 374 | 16.8 | 535 | 1 | ZN257_HUMAN |
| 16 | 373 | 16.7 | 470 | 1 | ZN436_HUMAN |
| 17 | 373 | 16.7 | 738 | 1 | ZN64_HUMAN |
| 18 | 373 | 16.7 | 821 | 1 | ZN64_HUMAN |
| 19 | 372 | 16.7 | 393 | 1 | ZN64_MOUSE |
| 20 | 367 | 16.5 | 818 | 1 | KR18_MOUSE |
| 21 | 364 | 16.3 | 428 | 1 | ZN83_HUMAN |
| 22 | 363.5 | 16.3 | 898 | 1 | ZN701_XENLA |
| 23 | 361 | 16.2 | 803 | 1 | ZN226_HUMAN |
| 24 | 360.5 | 16.2 | 630 | 1 | PRD5_HUMAN |
| 25 | 360 | 16.2 | 588 | 1 | ZN429_HUMAN |
| 26 | 359 | 16.1 | 1167 | 1 | ZN208_HUMAN |
| 27 | 357.5 | 16.0 | 739 | 1 | YD49_HUMAN |
| 28 | 355 | 15.9 | 538 | 1 | ZN155_HUMAN |
| 29 | 354 | 15.9 | 498 | 1 | ZN271_HUMAN |
| 30 | 354 | 15.9 | 498 | 1 | ZN234_HUMAN |
| 31 | 353 | 15.8 | 604 | 1 | ZN305_HUMAN |
| 32 | 351 | 15.7 | 474 | 1 | ZN141_HUMAN |
| 33 | 350.5 | 15.7 | 783 | 1 | ZN25_HUMAN |

| | | | | | | |
|----|-------|------|------|---|-------------|---------------------|
| 34 | 350 | 15.7 | 670 | 1 | ZN16_HUMAN | P17020 homo sapien |
| 35 | 350 | 15.7 | 706 | 1 | ZN225_HUMAN | Q9UK10 homo sapien |
| 36 | 349 | 15.7 | 489 | 1 | ZN271_HUMAN | Q9NG28 homo sapien |
| 37 | 348.5 | 15.6 | 439 | 1 | ZN208_XENLA | P18747 xenopus lae |
| 38 | 348.5 | 15.6 | 456 | 1 | ZN25_HUMAN | Q9U159 homo sapien |
| 39 | 348.5 | 15.6 | 606 | 1 | ZN214_HUMAN | P16374 mus musculus |
| 40 | 348.5 | 15.6 | 707 | 1 | ZN260_MOUSE | Q9UK13 homo sapien |
| 41 | 348 | 15.6 | 617 | 1 | ZN221_HUMAN | Q14709 homo sapien |
| 42 | 347 | 15.6 | 1029 | 1 | ZN197_HUMAN | Q9U1d9 homo sapien |
| 43 | 346 | 15.5 | 623 | 1 | ZN255_HUMAN | P34695 rana pipien |
| 44 | 345.5 | 15.5 | 335 | 1 | TF3A_RANPI | P17027 homo sapien |
| 45 | 345.5 | 15.5 | 643 | 1 | ZN23_HUMAN | |

ALIGNMENTS

RESULT 1
TF3A_YEAST
ID TF3A_YEAST STANDARD; PRT; 429 AA.
AC P39933;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Transcription factor TF3A (TF3A).
GN TF3A OR PZF1 OR TF3A1 OR YPR186C OR P9677.9.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92147684; PubMed=1737784;
RA Archambault J., Milne C.A., Schnappert K.T., Baum B., Friesen J.D.,
RA Segall J.E.
RT "The deduced sequence of the transcription factor TF3A from
RT Saccharomyces cerevisiae reveals extensive divergence from Xenopus
RT TF3A".
RL J. Biol. Chem. 267:3282-3288(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92237295; PubMed=1570325;
RA Woychik N.A., Young R.A.
RT "Genes encoding transcription factor TF3A and the RNA polymerase
RT common subunit RPB6 are divergently transcribed in Saccharomyces
RT cerevisiae".
RL Proc. Natl. Acad. Sci. U.S.A. 89:3999-4003(1992).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=6288C / AB972;
RC MEDLINE=97131271; PubMed=9169875;
RA Bussey H., Storms R.K., Ahmed A., Alberman K., Allen E., Ansoorge W.,
RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
RA Dietrich F.S., Delius H., DiPaolo T., Dubois E., Duesterhoeft A.,
RA Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
RA Hall J., Hebling M., Heumann K., Hilbert H., Hillier L.W.,
RA Hunnicke-Smith S., Hyman R., Johnston M., Kalman S., Klein K.,
RA Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,
RA Marthe R., Messenguy F., Mewes H.-W., Mitalpatis S., Moestl D.,
RA Mueller-Auer S., Namath A., Nentwich U., Oetner P., Pearson D.,
RA Petel F.X., Pohl T.M., Purnelle D., Schater M., Scharer M.,
RA Scherens B., Schramm S., Schroeder W., Sidic A.M., Yetteil H.,
RA Uristetzer L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,
RA Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,
RA Zhong W.W., Zollner A., Vo D.H., Han J.
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI";
RL Nature 387:103-105(1997)
CC - FUNCTION: INTERACTS WITH THE INTERNAL CONTROL REGION (ICR) OF
CC APPROXIMATELY 50 BASES WITHIN THE 5S RNA GENES. IS REQUIRED FOR
CC CORRECT TRANSCRIPTION OF THESE GENES BY RNA POLYMERASE III. ALSO
CC BINDS THE TRANSCRIBED 5S RNA S.

```

CC -1- SUBCELLULAR LOCATION: Nuclear.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M80611; AAB08014.1; -.
CC EMBL: M96638; -, NOT ANNOTATED_CDS.
CC EMBL: U25841; AAB64615.1; -.
CC PIR: S20050; S20050.
CC Germonline; 144451; -.
CC TRANSFAC; T03530; -.
CC SGD; S0006390; PZFL.
CC GO; GO:0005667; C:transcription factor complex; IPI.
CC GO; GO:0003709; F:RNA polymerase III transcription factor act. .; IPI.
CC GO; GO:0006384; P:transcription initiation from Pol III promoter; IPI.
CC InterPro; IPR007087; Znf_C2H2.
CC SMART; SM00355; Znf_C2H2; 9.
CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 6.
CC PROSITE; PS00157; ZINC_FINGER_C2H2_2; 6.
CC Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
CC RNA-binding; Repeat; Nuclear protein.
CC DOMAIN 24 43 SER-RICH.
CC ZN_FING 49 74 C2H2-TYPE.
CC ZN_FING 80 102 C2H2-TYPE.
CC ZN_FING 108 130 C2H2-TYPE.
CC ZN_FING 134 159 C2H2-TYPE.
CC ZN_FING 163 186 C2H2-TYPE.
CC ZN_FING 194 219 C2H2-TYPE.
CC ZN_FING 222 244 C2H2-TYPE.
CC ZN_FING 253 277 C2H2-TYPE.
CC DOMAIN 321 327 ARG/LYS-RICH (BASIC).
CC ZN_FING 365 389 C2H2-TYPE.
CC SEQUENCE 429 AA; 50027 MW; 209B1ED6A20422D9 CRC64;
SQ
Query March 37.6%; Score 837.5; DB 1; Length 429;
Best local similarity 43.7%; Pred. No. 1.3e-42;
Matches 174; Conservative 62; Mismatches 135; Indels 27; Gaps 11;
CC 1 MSEDDETKSISLIS--SSSSSPKXYCTYEGCDKAYNPRLGFQ--LRTHSNDPXYK 57
CC 23 ISRSBSSSLNLSINSTRSSSNRPXYPCDYDGDCKAFTRPSILTEHQLSHVQGLRQOC 82
CC 58 TVDDCDKAFPKKSHLEHTIVHSEKKPRHCVGCGVNSRQHLKRMHETHTKSPKCTPEN 117
CC 83 --DKCASFVKSHLERHLVTHSDTKPRQCYCGAGVTRQOLKHEVTHKSFICPREG 140
CC 118 COEAFYKQSLRHHLISVHEKTLTCKQCNKVFTPRSKLAQHLKHHGG--SPAYQCDHPG 175
CC 141 CNLRPKRPPQRLRAHLISVHLKLTCPHCKNSPQRYRLRNHLISKHDEVEHPYCTPAG 200
CC 176 CFKNQFVSVLQPHIKQSHPKLCPKCGKCGYKGLSSHMLSHDSDTMIKITWDCYDV 235
CC 201 CCKEPRISQLOSHIKNDHPKLCPICKPCVGENGLQMHMLIHDSLVTKMKWCKHICPD 260
CC 236 GKPAKKNLVHYNVFN--DGNIPDGL--LKEPEVKKLEMLLDGSGKLNNLHELETKLKY 293
CC 261 MSFSKRDHLTHVSGIHTEEDIPLEKTKYISDIQGL--VQDHGVQGLN-----SKHSE 312
CC 294 EDEEEDSDIDKRSDDVSQMSAQRISKFTASLE--GSKSVSKLISNGKKINCPKNC 352
CC 313 QDEKISRLKRRKRLKLTENNVEFQNEVDLEKRLIESGNGNLNLLNTVGRKYRCFYNNC 372
CC 353 DRMFSEYDRLRHL--KMHNDLQRISEFLNISTEKER 387
CC 373 SRTFKTEKYEKHIIDKHVKHLEKTLIQ-----EKEE 404

```

```

RESULT 2
ID TF3A_SCHPO STANDARD; PRT; 374 AA.
AC Q9UTD5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Transcription factor IIIA (Factor A) (TFIIIA).
GN SFC2 OR SPAC144.09C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A., AND FUNCTION.
RX MEDLINE=22082319; PubMed=12087160;
RA Schullman D.B., Setzer D.R.;
RT "Identification and characterization of transcription factor IIIA from
RT Schizosaccharomyces pombe."
RL Nucleic Acids Res. 30:2772-2781(2002).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Woods J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Murphy K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares K., Squires S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Voicakeert G., Aert R., Robben J., Grymopiez B.,
RA Weltjens I., Vansteegle E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambut R., Fumelle B.,
RA Goffeau A., Cadieu E., Driano S., Gloux S., Leleau V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hunt S.M.,
RA Lucas M., Roeder M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Useery D., Bartell B.G., Nurre P.;
RT "The genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880(2002).
-1- FUNCTION: Is required for correct transcription of 5S RNA genes by
RNA polymerase III. Also binds the transcribed 5S RNA's. Initiates
transcription of the 5S ribosomal RNA gene.
-1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AY091590; AAM00046.1; -.
CC EMBL: AL132675; CAB56889.1; -.
CC PIR: T37676; T37676.
CC HSSP; P07248; 1ARD.
CC GeneDB; Spombe; SPAC144.09c; -.
CC InterPro; IPR007087; Znf_C2H2; 10.
CC Pfam; PF00096; Zf-C2H2; 10.
CC ProDom; PD000003; Znf_C2H2; 1.
CC SMART; SM00355; Znf_C2H2; 9.

```

DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 8.
 DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 9.
 KM Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
 FW RNA-binding; Repeat; Nuclear protein.
 FT ZN_FING 23 47 C2H2-TYPE.
 FT ZN_FING 53 77 C2H2-TYPE.
 FT ZN_FING 83 107 C2H2-TYPE.
 FT ZN_FING 113 138 C2H2-TYPE.
 FT ZN_FING 144 169 C2H2-TYPE.
 FT ZN_FING 204 226 C2H2-TYPE.
 FT ZN_FING 236 261 C2H2-TYPE.
 FT ZN_FING 267 291 C2H2-TYPE (ATYPICAL).
 FT ZN_FING 349 374 C2H2-TYPE.
 SQ SEQUENCE 374 AA; 43851 MW; 7469C701FF08FF6 CRC64;
 Query Match 22.2%; Score 495; DB 1; Length 374;
 Best Local Similarity 32.8%; Pred. No. 1.6e-22;
 Matches 123; Conservative 58; Mismatches 160; Indels 34; Gaps 13;

QY 15 SSSSSSRPKYITTYGSCDKANRPSLBQHLRTHSNDRPYKCTVDCCDAFFPKSHLET 74
 DB 13 SKNIRSAKIFHCPCYECCGKYSRPSLLEQHLRTHSNERPVCVYGCXKAFYKSHLKI 72
 QY 75 HIVESEKPEFGCV--CGKVNRSQHLKHEITH--TSEFKTFENCGAIFYKQSLDH 130
 DB 73 HRCCHTNVAFPCFCHYGCAQFYTOQLEHRIEVRKRPYACTWEGCECFKHQQLRS 132
 QY 131 HILSVHEKTL---TKQCNKVFTRPSKLAQHLKHHGSPAYQCDHPGC--FKNFQTS 184
 DB 133 HICAGTHLLPYRCTYQDCERFATKQKQNHVNRHAKETISCSHESCVGHEGEXMS 192
 QY 185 VLOFHKKQSHPKKCKKCGKCGKGLSSHMLSHDSTMIKWTQCYDV---GKFAK 240
 DB 193 QLONHIREAHVP--SCSICGRQFKTAHLRHHVTLH--QTTLERKTYHCPMEGCKKSFT 249
 QY 241 KQELVEHYVIFPDGNP--DILLKETEYK-KLENLLDQSGSLNNHLETERKAYEER 296
 DB 250 SSALKKHISVIEGNAFHCHDCGTFYGYKMLQRLHREGT-CKAKHKYINECGIKHDG 308
 QY 297 EDEEDSLDERKSDVRSDSMS--AQRISIKSFASLESGKSVSKLISNGKINCPKNNCDR 354
 DB 309 IGVVAIHDCKEKELSNLVDAKTIINEVTG--HGYSK-----EAREYSCSPEECNY 358
 QY 355 MSREYDLRRLKWH 369
 DB 359 RPKRLYDMRHLNSH 373

DR ZN43_HUMAN STANDARD; PRT; 803 AA.
 AC P17038; P28160; O96DG1;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Zinc finger protein 43 (Zinc protein HTF6) (Zinc finger protein K0X27).
 GN ZNF43 OR ZNF39 OR K0X27.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxId:9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=T-cell;
 RX MEDLINE=91279444; PubMed=1711675;
 RA Lowering R., Trowsdale J.;
 RA "A gene encoding 22 highly related zinc fingers is expressed in
 RL lymphoid cell lines.";
 RN Nucleic Acids Res. 19:2921-2927 (1991).
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lymph;

RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Cassavani T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loughell N.A., Peters G.J., Abramson R.D., Muliyil S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Guntarone P.H.,
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman A.J., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smolins D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN (3)
 RP SEQUENCE OF 38-190 FROM N.A.
 RX MEDLINE=91219421; PubMed=2023909;
 RA Bellefroid E.J., Poncelet D.A., Lecocq P.J., Revelant O.,
 RA Marciel J.A.;
 RT "The evolutionarily conserved Kruppel-associated box domain defines a
 RT subfamily of eukaryotic multifingered proteins.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:3608-3612 (1991).
 RN (4)
 RP SEQUENCE OF 476-531 FROM N.A.
 RC TISSUE=Lymphoid;
 RX MEDLINE=91145339; PubMed=2288909;
 RA Thiesen H.-J.;
 RT "Multiple genes encoding zinc finger domains are expressed in human T
 RT cells.";
 RL New Biol. 2:363-374 (1990).
 CC -1- FUNCTION: May be involved in transcriptional regulation.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- TISSUE SPECIFICITY: T AND B CELL LINES.
 CC -1- SIMILARITY: BELONGS TO THE KRUEPPL FAMILY OF C2H2-TYPE ZINC-
 CC FINGER PROTEINS.
 CC -1- SIMILARITY: Contains 1 KRAB domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib.ch).
 CC -----
 DR EMBL, X59244; CAA41932.1; -;
 DR EMBL, BC006528; AAH06528.1; -;
 DR EMBL, M61869; AAA58674.1; -;
 DR EMBL, X52358; CAA36584.1; -;
 DR PIR, S26823; S26823.
 DR HSSP: P08048; 7ZNF.
 DR TRANSFAC, T04986; -;
 DR GeneW, HGNC:13109; ZNF43.
 DR MIM, 603972; -;
 DR GO, GO:0003677; F:DNA binding; TAS.
 DR InterPro, IPR001909; KRAB.
 DR InterPro, IPR007087; Znf_C2H2.
 DR InterPro, IPR007086; Znf_C2H2_sub.
 DR Pfam, PF01352; KRAB; 1.
 DR Pfam, PF00096; zf-C2H2; 21.
 DR PRINTS, PR00048; ZINC_FINGER.
 DR ProDom, PD000003; Znf_C2H2; 16.
 DR SMART, SM00349; KRAB; 1.
 DR SMART, SM00355; Znf_C2H2; 22.
 DR PROSITE, PS00805; KRAB; 1.
 DR PROSITE, PS00028; ZINC_FINGER_C2H2_1; 19.

DR PROSITE: PS0157; ZINC_FINGER C2H2_2; 22.
 KW Transcription regulation; DNA-binding; zinc-finger; Metal-binding;
 NW Nuclear protein; Repeat.

FT ZN_FING 168 190 KRAB.
 FT ZN_FING 196 218 C2H2-TYPE. (DEGENERATE).
 FT ZN_FING 224 246 C2H2-TYPE (DEGENERATE).
 FT ZN_FING 252 274 C2H2-TYPE (DEGENERATE).
 FT ZN_FING 280 302 C2H2-TYPE.
 FT ZN_FING 308 330 C2H2-TYPE.
 FT ZN_FING 336 358 C2H2-TYPE.
 FT ZN_FING 364 386 C2H2-TYPE.
 FT ZN_FING 392 414 C2H2-TYPE.
 FT ZN_FING 420 442 C2H2-TYPE.
 FT ZN_FING 448 470 C2H2-TYPE.
 FT ZN_FING 476 498 C2H2-TYPE.
 FT ZN_FING 504 526 C2H2-TYPE.
 FT ZN_FING 532 554 C2H2-TYPE.
 FT ZN_FING 560 582 C2H2-TYPE.
 FT ZN_FING 588 610 C2H2-TYPE.
 FT ZN_FING 616 638 C2H2-TYPE.
 FT ZN_FING 644 666 C2H2-TYPE.
 FT ZN_FING 672 694 C2H2-TYPE.
 FT ZN_FING 700 722 C2H2-TYPE.
 FT ZN_FING 728 750 C2H2-TYPE.
 FT ZN_FING 756 778 C2H2-TYPE.
 FT CONFLICT 181 181 P -> S (IN REF. 3).
 FT CONFLICT 712 712 P -> S (IN REF. 2).
 SQ SEQUENCE 803 AA; 93487 MW; 88C41B44DFE2533 CRC64;

Query Match 17.9%; Score 400; DB 1; Length 803;
 Best Local Similarity 31.9%; Pred. No. 1,4e-16;
 Matches 121; Conservative 53; Mismatches 171; Indels 34; Gaps 19;

QY 2 SESDETYSISLSSSS--SRPKYICTYEGCDKAVNRPDLRQHLPRTSNDPRYCTV 59
 DB 367 TEGCEAFSRSSNLTGKKIKHTKKPKYC--EECGKAFKMSKLTGKLTGKEPKYC-- 422
 QY 60 DDCDKAFRRKSHLTHIVSHSEKKEPHGCVGKGVNRQHLKREIHT--KSKCTFEN 117
 DB 423 EECGKAFNWPSTLTGKRIHITGKEPKYCEGKAFNPSNLTGKRIHTGKPKYC--EE 480
 QY 118 COEAFYGHQST--RHILSVHEKTLTCKCNKVFTRPRLAQHKLKHGSPAYOCDDPGC 176
 DB 481 CGKAFSSSNLTGKRIHITGKEPKYCEGKAFKMSKLTGKLTGKEPKYC--YCCEB--C 537
 QY 177 FKQFQTVSVLQFH--IKQSHPKLCKPKCGKGVGKGLSSHMLSHDSTMIKTIWTCYCD 234
 DB 538 GKAFNHFSILTKHRIHTGKEPKYCEGKAFQSSNLTGKRIH--TGKFKYCEB--C 593
 QY 235 VGR-FAKKNELVEHNYTFHONIPDDLKTEVEYKLEMLDQGSKLNNLHELETKLVE 293
 DB 594 -GKAFQSSNLTGKRIHITGKEPKYC--YKCEGKAFNPSNLTGKRIH--TEKPKYCE 647
 QY 294 E-DEDEDESDLEKRSVDS--MSAQRISKSTASLEGSYSKLSNKGKINCPCN 350
 DB 648 ECGKAFKMSSTLTGKRIHITGKEPKYCEGKAF--KLSSTLTGKRIHITGKEPKYCEK- 704
 QY 351 NCDRMPSEYDLRRHLKWH 369
 DB 705 -CGKAFNRPSNLTGKRIH 722

RESULT 4
 Z492 HUMAN STANDARD; PRT; 574 AA.
 AC Q9255;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Zinc finger protein 492 (Fragment).
 GN ZNF492 OR KIA1473.
 OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=20277482; PubMed=10819311;
 RA Nagase T., Kikuno R., Ishikawa K.-I., Hirosewa M., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XVII.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 7:143-150(2000).
 CC -1- FUNCTION: May function as a transcription factor.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
 CC FINGER PROTEINS.
 CC -1- SIMILARITY: Contains 1 KRAB domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: AB040906; BAA9597.1; -
 CC HSP: P07248; 1ARD.
 CC Genew: HGNC:23707; ZNF492.
 CC Interpro: IPR001909; KRAB.
 CC Interpro: IPR007087; Znf_C2H2.
 CC Interpro: IPR007086; Znf_C2H2_sub.
 CC Pfam: PF01352; KRAB; 1.
 CC Pfam: PF00906; zf-C2H2; 13.
 CC PRINTS: PR00046; ZINCFINGER.
 CC PRODOM: PD000003; Znf_C2H2; 2.
 CC SMART: SM00349; KRAB; 1.
 CC SMART: SM00355; Znf_C2H2; 13.
 CC PROSITE: PS00805; KRAB; 1.
 CC PROSITE: PS00028; ZINC_FINGER C2H2_1; 12.
 CC PROSITE: PS00157; ZINC_FINGER C2H2_2; 13.
 CC Hypothetical protein; Transcription regulation; DNA-binding;
 CC Zinc-finger; Metal-binding; Nuclear protein; Repeat.
 FT ZN_FING 1 1 KRAB.
 FT ZN_FING 15 86 C2H2-TYPE.
 FT ZN_FING 184 206 C2H2-TYPE.
 FT ZN_FING 212 234 C2H2-TYPE.
 FT ZN_FING 240 262 C2H2-TYPE.
 FT ZN_FING 268 290 C2H2-TYPE.
 FT ZN_FING 296 318 C2H2-TYPE.
 FT ZN_FING 324 346 C2H2-TYPE.
 FT ZN_FING 352 374 C2H2-TYPE.
 FT ZN_FING 380 402 C2H2-TYPE.
 FT ZN_FING 408 430 C2H2-TYPE.
 FT ZN_FING 436 458 C2H2-TYPE.
 FT ZN_FING 464 486 C2H2-TYPE.
 FT ZN_FING 492 514 C2H2-TYPE (DEGENERATE).
 FT ZN_FING 520 542 C2H2-TYPE.
 SQ SEQUENCE 574 AA; 65951 MW; 482B5F1176E0A5B0 CRC64;

Query Match 17.9%; Score 398; DB 1; Length 574;
 Best Local Similarity 30.2%; Pred. No. 1.3e-16;
 Matches 121; Conservative 48; Mismatches 167; Indels 64; Gaps 18;

QY 20 SRPKYICTYEGCDKAVNRPDLRQHLPRTSNDPRYCTVDDCKAFRRKSHLTHIVSH 79
 DB 207 GSEKPKYC--KECGKAFNPSNLTGKRIHTGKPKYC--EECGKAFRRKSHLTHIKTIH 262
 QY 80 SEKKPHGCVGKGVNRQHLKREIHT--KSKCTFENCOEAFYGHQST--RHILSVH 136
 DB 263 TGKPKYCEGKAFNPSNLTGKRIHTGKPKYC--EECGKAFSSNLTGKRIHAG 320
 QY 137 EKLTLCKQCNKVFTRPRLAQHKLKHGSPAYOCDDPGCCKRNFQTVSVLQFH--IKQSH 194

```

Db      321 EKYKCECGCAKASQSTLTTHKI -IHTGEKYKCE- -CGKAFSQSLTTHKRIHSGE 377
QY      195 PKUCPCPCGKGCCKKGLSSHMLSHDSTMIKTIWTCYDVGFAKKNELVEHYNI PHDG 254
Db      378 KPYKCECGCAKAFQSSLTTHKRIHAGE -KPYKCEVCCKA -FSRSHLTTHKRI HTG 432
QY      255 NIPDDILKETEVKLEMLLQGSKANLHLETEKLVKEEDEDEDSLDKRSVDYRSDS 314
Db      433 EKP---YKCECGCAKAFNLSSQLTTHKLIHTGE -KPYKCEBC-----GKAFQSSST 478
QY      315 MSAQST-----KSPFASLEGSKSVSKLISNGKINCPKNNCDMPFSEYD 361
Db      479 LSHKVIHTGEKPYKCECGKAFNQ- -SHLTTHKMIHTGEKPYKC- -ECCGKAFNNSS1 534
QY      362 LRRHLKWHD-----DNLQRIEPLNLSIEKEE 387
Db      535 LNHKMIHTGEKLYKPESCNACDNIAKISKYRNCAGEK 574

RESULT 5
TF3A_HUMAN          STANDARD;          PRT;          423 AA.
ID   TF3A_HUMAN          STANDARD;          PRT;          423 AA.
AC   092664; O12963; Q13097;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Transcription factor IIAA (Factor A) (TFIIIA).
GN   GTF3A.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX   Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Fetal brain;
RX   MEDLINE=95309028; PubMed=7789179;
RA   Arakawa H., Nagase H., Hayashi N., Ogawa M., Nagata M.,
RT   Fujimura T., Takahashi E., Shin S., Nakamura Y.;
RT   "Molecular cloning, characterization, and chromosomal mapping of a
RT   novel human gene (GTF3A) that is highly homologous to Xenopus
RT   transcription factor IIAA."
RT   Cytogenet. Cell Genet. 70:235-238 (1995).
RN   [2]
RP   SEQUENCE OF 61-423 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=95347600; PubMed=7622052;
RA   Drew P.D., Nagle J.W., Canning R.D., Ozato K., Biddison W.E.,
RT   Becker K.G.;
RT   "Cloning and expression analysis of a human cDNA homologous to
RT   xenopus TFIIIA."
RT   Gene 159:215-218 (1995).
RN   [3]
RP   CHARACTERIZATION.
RX   MEDLINE=94342241; PubMed=8063702;
RA   Moorefield B., Roeder R.G.;
RT   "Purification and characterization of human transcription factor
RT   IIAA."
RT   J. Biol. Chem. 269:20857-20865 (1994).
CC   -I- FUNCTION: INTERACTS WITH THE INTERNAL CONTROL REGION (ICR) OF
CC   APPROXIMATELY 50 BASES WITHIN THE 5S RNA GENES, IS REQUIRED FOR
CC   CORRECT TRANSCRIPTION OF THESE GENES BY RNA POLYMERASE III. ALSO
CC   BINDS THE TRANSCRIBED 5S RNA'S. MAY INITIATE TRANSCRIPTION OF THE
CC   5S RIBOSOMAL RNA GENE AND MAINTAIN THE STABILITY OF TRANSCRIPTION
CC   OF OTHER GENES.
CC   -I- SUBCELLULAR LOCATION: Nuclear.
CC   -I- TISSUE SPECIFICITY: Ubiquitous.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial

```

```

CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
Db      EMBL; D12257; AAA06988.1; -
DR      EMBL; U20272; AAA5623.1; -
DR      EMBL; U14134; AAA21873.1; -
DR      HSSP; P03001; 1TF3.
DR      TRANSFAC; T04953; -
DR      Genew; HGNC:4662; GTF3A.
DR      MTM; 600860; -
DR      GO; GO:0003709; F:RNA polymerase III transcription factor act. .; TAS.
DR      GO; GO:0009303; P:RNA transcription; TAS.
DR      GO; GO:0006383; P:transcription from Pol III promoter; TAS.
DR      InterPro; IPR007087; Znf C2H2.
DR      Pfam; PF00096; zf-C2H2; 9.
DR      SMART; SM00355; Znf_C2H2; 9.
DR      PROSITE; PS00028; ZINC_FINGER_C2H2_1; 7.
DR      PROSITE; PS0157; ZINC_FINGER_C2H2_2; 8.
KW      Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
KW      RNA-binding; Repeat; Nuclear protein; Polymorphism.
FT      ZN_FING          98      122
FT      ZN_FING          128     152
FT      ZN_FING          158     183
FT      ZN_FING          190     212
FT      ZN_FING          220     244
FT      ZN_FING          247     271
FT      ZN_FING          275     297
FT      ZN_FING          304     329
FT      ZN_FING          335     359
FT      VARIANT          303
FT      CONFLICT          214     214
FT      CONFLICT          221     221
FT      CONFLICT          223     248
FT      CONFLICT          378     387
SQ      SEQUENCE          423 AA; 46847 MW; A627D064A43FB6F0 CRC64;
Query Match          17.8%; Score 397.5; DB 1; Length 423;
Best Local Similarity 31.1%; Pred. No. 9,9e-17;
Matches          99; Conservative          55; Mismatches          95; Indels          69; Gaps          16;
QY      8 KSISSL-----ISSSSSR-----PKYICYEGGCKAYNRPSLLEOHLRTSHNDR 53
Db      67 ESVSLLTADAFIAGESSAPPPRPALPRFICSPPOCSANYSKAMKLDHLCCKTGRR 126
QY      54 PKCTVDDCDKAFKFRKSHLETHIVSHSEKKEPHCSV--CGKGVNSKQHLKRR-ETTH--- 107
Db      127 PFCVCEGCGKAFIDYHLSRHLTHTEGKPFVCANAGCDQKFNKSNLKKHFERKHENQ 186
QY      108 TKSFKTEFENCOEAFYKHQSIR-HHILSYHEKTLVCKQ--CNKVFTRPSKLAQHLKHNG 164
Db      187 QKQYICSPEDCKCKTKRKHQQLKHQCCQNNNELPFKCTQDGCCKGHRPSKLRHAKAHBG 246
QY      165 GSPAYQCHPGCEFKNPFQMSVLQFHIKOSHPR----- 196
Db      247 ----YVC-QKGSFPAKWTLELLKVRHREHKELLCEVCRKTFKRKDYLKQMKTHAPER 301
QY      197 --LKCEK--CGKGCYKGGKGLSHMLS-HDSTMIKTIWTCYDVGK-FKKKELVEHYNI 250
Db      302 DVCRCPRGCGRTYTTVNLQSHLISFHEE--RPVCEHAGCGCTFAMKQSLTRH-AV 357
QY      251 FHDGNIIPDDLKETEYVK 268
Db      358 VHD---PDKKKKLVKK 372

RESULT 6
ZN85_HUMAN          STANDARD;          PRT;          595 AA.
ID   ZN85_HUMAN          STANDARD;          PRT;          595 AA.
AC   003923;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)

```

DE Zinc finger protein 85 (Zinc finger protein HPF4) (HTEFL).
GN ZNF85.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN (1)
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC TISSUE=Placenta;
RA MEDLINE=99053537; PubMed=9839802;
RA Poncelet D.A., Bellefroid E.J., Bastiaens P.V., Demotie M.A.,
RA Muller M., Marital J.A.; Alami Y., Devos N., Lecocq P.J., Ogawa T.,
RT "Functional analysis of ZNF85 KRAB zinc finger protein, a member of
RT the highly homologous ZNF91 family.";
RL DNA Cell Biol. 17:931-943(1998).
RN (2)
RP SEQUENCE OF 1-196 FROM N.A.
RX MEDLINE=91219421; PubMed=2023909;
RX Bellefroid E.J., Poncelet D.A., Lecocq P.J., Revelant O.,
RA Marital J.A.;
RT "The evolutionarily conserved Kruppel-associated box domain defines a
RT subfamily of eukaryotic multifingered proteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:3608-3612(1991).
CC -1- FUNCTION: Transcriptional repressor.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: Expressed preferentially in testicular
CC tissues.
CC -1- DEVELOPMENTAL STAGE: Expressed early during embryonic development.
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -1- SIMILARITY: Contains 1 KRAB domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U35376; AAA79179.1; -;
DR EMBL: M61866; AAA52689.1; -;
DR EMBL: M61868; AAA56671.1; -;
DR PIR: C39384; C39384.
DR PIR: G02075; G02075.
DR HSSP: P08048; ZNF.
DR TRNSPAC: T04990; -;
DR Genew: HGNC:13160; ZNF85.
DR MIT: 603899; -;
DR GO: GO:0005634; C:nucleus; TAS.
DR GO: GO:0003714; F:transcription co-repressor activity; TAS.
DR GO: GO:0003700; F:transcription factor activity; TAS.
DR InterPro: IPR001909; KRAB.
DR InterPro: IPR007087; Znf_C2H2.
DR InterPro: IPR007086; Znf_C2H2_sub.
DR Pfam: PF01352; KRAB. 1.
DR Pfam: PF00096; ZF-C2H2. 15.
DR PRINTS: PR00048; ZINC_FINGER.
DR ProDom: PD000003; Znf_C2H2. 13.
DR SMART: SM00349; KRAB. 1.
DR SMART: SM00355; Znf_C2H2. 15.
DR PROSITE: PS00805; KRAB. 1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1. 14.
DR PROSITE: PS00157; ZINC_FINGER_C2H2_2. 15.
DR Transcription regulation; Zinc-finger; DNA-binding; Metal-binding;
KW Nuclear protein; Repeat; Repressor.
FT DOMAIN 4 75 KRAB
FT ZN_FING 146 168 C2H2-TYPE (DEGENERATE).
FT ZN_FING 174 196 C2H2-TYPE.
FT ZN_FING 202 224 C2H2-TYPE.
FT ZN_FING 230 252 C2H2-TYPE.
FT ZN_FING 258 280 C2H2-TYPE.

FT ZN_FING 286 308 C2H2-TYPE.
FT ZN_FING 314 336 C2H2-TYPE.
FT ZN_FING 342 364 C2H2-TYPE.
FT ZN_FING 370 392 C2H2-TYPE.
FT ZN_FING 398 420 C2H2-TYPE.
FT ZN_FING 426 448 C2H2-TYPE (DEGENERATE).
FT ZN_FING 454 476 C2H2-TYPE.
FT ZN_FING 482 504 C2H2-TYPE.
FT ZN_FING 510 532 C2H2-TYPE.
FT ZN_FING 538 560 C2H2-TYPE.
FT ZN_FING 566 588 C2H2-TYPE.
FT CONFLICT 84 84 R -> Q (IN REF. 2; AAA52689).
FT CONFLICT 115 115 R -> I (IN REF. 2; AAA58671).
FT CONFLICT 177 177 T -> R (IN REF. 2; AAA58671).
FT CONFLICT 184 184 G -> R (IN REF. 2; AAA58671).
SQ SEQUENCE 595 AA; 68718 MW; 44AA0A236D2D43B CRC64;
Query Match 17.8%; Score 396.5; DB 1; Length 595;
Best local similarity 31.7%; Pred. No. 1,6e-16;
Matches 120; Conservative 48; Mismatches 160; Indels 51; Gaps 17;
QY 10 ISSLISSSS-SRPKXYICTYEGCDKAYNRPSLLBQHRTSHNDRPYKCTVDDCKAFPR 68
DB 186 ISCLTSHSRITHTVNFYKC--EECGKAFWNSLTLYHKRLHTGEKPYKC--EECGKAFNQ 241
QY 69 KSHLETHIVSHSEKPKFPCVCGKGVNSROHLKRHEITH--KSFCKTFENCORAFYHQ 126
DB 242 SSNLIHKHKTHTGEKPYKCECGKAFRPSLTLYHKRLHTGEKPYKC--KCGKAFRRSS 299
QY 127 SL-RHHLIVSHEKTLTCKQKNVFTPRSKLAQHKLKHGSSPAYQCDHPGCFKNFGQWSV 185
DB 300 TLTHRKHTHGEKPYKCECGKAFQSSNLTLYHKRLHTGEKPYKCK--GKAFNQSAAH 356
QY 186 LQHH--IKQSHPKLCKPGKCGVGGKGLSSHLSDHSDMIKLTQDVCYGFARKNE 243
DB 357 LTHLEVHTGEKPYKCECGKAFNHSHTLYHKRLHT--TGEKPYKCKEC--GKAFRRSS 411
QY 244 LVEHYNIFFHGNIPDDLKTEVYKLEMLDQSKLNNLHELETREKLVDEDEDEDSLT 303
DB 412 TLTHKRLHTHGE-----KPKYKCECGKAFNQSXL-----TEHKRLHTGEKPYE--- 455
QY 304 DEKRSIVRSQMSAQRISKFTFASL-----EGSK-----SVSKLSINSKGLNCPKN 350
DB 456 CEKCGKAFNQSNTLTHKSHTEKPYKCECGKAFWPSLTLYHKRLHTGEKPYKC--E 513
QY 351 NCDRMFSREYDARHKLKQH 369
DB 514 ECGKAFNQSRLTHKRLH 532
RESULT 7
ZN93 HUMAN STANDARD; PRT; 616 AA.
ID ZN93 HUMAN
AC P35789; Q9Y2N8;
DT 01-JUN-1994 (Rel. 29, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Zinc finger protein 93 (Zinc finger protein HTP34) (Fragment).
GN ZNF93.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN (1)
RP SEQUENCE FROM N.A.
RA Lamerdin J.E., McCreedy P.M., Skowronski E., Sakaidais G.,
RA Burkhardt-Schultz K., Gordon L., Dias J., Scott D., Stiliwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Dangnan L.,
RA Erler A., Christensen M., Georgescu A., Avila J., Attix C.,
RA Andrade T., Trankleim M., Amico-Keller G., Coefield J., Duarte S.,
RA Lucas S., Bruce R., Thomas P., Quan G., Krommiller B., Arellano A.,
RA Sanders C., Ow D., Nolan M., Trong S., Kobayashi A., Olsen A.S.,
RA Carrano A.V.;

```

RT "Sequence analysis of a 5.7 Mb region in 19p13.1 between OLFR and
RT D19S455."
RT Submitted (APR-1999) to the EMBL/GenBank/DBJ databasses.
RN [2]
RP SEQUENCE OF 1-191 FROM N.A.
RX MEDLINE=91219421; PubMed=2023909;
RA Bellefroid B.J., Poncelet D.A., Lecocq P.J., Revelant O.,
RA Martial J.A.;
RT "The evolutionarily conserved Kruppel-associated box domain defines a
RT subfamily of eukaryotic multifingered proteins.";
RT Proc. Natl. Acad. Sci. U.S.A. 88:3608-3612(1991).
CC CC -1- FUNCTION: May be involved in transcriptional regulation.
CC CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC CC -1- DEVELOPMENTAL STAGE: Expressed early during embryonic development.
CC CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC CC FINGER PROTEINS.
CC CC -1- SIMILARITY: Contains 1 KRAB domain.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (see http://www.isb.ch/announce/
CC CC or send an email to license@isb.ch).
CC CC -----
CC CC EMBL, AC007204; AAD22981.1; -.
CC CC HSPB, M61873; AAA83548.1; -.
CC CC HSSP; P08046; IAH1.
CC CC Genew, HGNC:13169; ZNF93.
CC CC MIM, 603975; -.
CC CC DR GO; GO:0005634; C:nucleus; NAS.
CC CC DR GO; GO:0003700; F:transcription factor activity; NAS.
CC CC DR GO; GO:0008270; F:zinc ion binding; NAS.
CC CC DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
CC CC DR InterPro; IPR001087; Znf_C2H2.
CC CC DR InterPro; IPR007086; Znf_C2H2_sub.
CC CC DR Pfam; PF01352; KRAB; 1.
CC CC DR Pfam; PF00096; zf-C2H2; 16.
CC CC DR PRINTS; PR00048; ZINCINGER.
CC CC DR Prodom; PD000003; Znf_C2H2; 7.
CC CC DR SMART; SM00349; KRAB; 1.
CC CC DR SMART; SM00355; Znf_C2H2; 16.
CC CC DR PROSITE; PS50085; KRAB; 1.
CC CC DR PROSITE; PS50029; ZINC_FINGER_C2H2_1; 16.
CC CC DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 17.
CC CC K4 Transcription regulation; Zinc-finger; DNA-binding; Metal-binding;
CC CC Nuclear protein; Repeat.
CC CC FT NON TER 1 71
CC CC FT DOMAIN <1 71
CC CC FT ZN_FING 141 163 KRAB.
CC CC FT ZN_FING 169 191 C2H2-TYPE.
CC CC FT ZN_FING 197 219 C2H2-TYPE.
CC CC FT ZN_FING 225 247 C2H2-TYPE.
CC CC FT ZN_FING 253 275 C2H2-TYPE.
CC CC FT ZN_FING 281 303 C2H2-TYPE.
CC CC FT ZN_FING 309 331 C2H2-TYPE.
CC CC FT ZN_FING 337 359 C2H2-TYPE.
CC CC FT ZN_FING 365 387 C2H2-TYPE.
CC CC FT ZN_FING 393 415 C2H2-TYPE.
CC CC FT ZN_FING 421 443 C2H2-TYPE.
CC CC FT ZN_FING 449 471 C2H2-TYPE.
CC CC FT ZN_FING 477 499 C2H2-TYPE.
CC CC FT ZN_FING 505 527 C2H2-TYPE.
CC CC FT ZN_FING 533 555 C2H2-TYPE.
CC CC FT ZN_FING 561 583 C2H2-TYPE.
CC CC FT ZN_FING 589 611 C2H2-TYPE.
CC CC FT ZN_FING 71 71 S -> SGP (IN REF. 2).
CC CC FT CONFLICT 85 85 Q -> H (IN REF. 2).
CC CC FT CONFLICT 115 115 S -> R (IN REF. 2).
CC CC FT CONFLICT 132 132 C -> S (IN REF. 2).
CC CC FT CONFLICT 149 149 V -> D (IN REF. 2).
CC CC FT CONFLICT 149 149

```

| SEQ | SEQUENCE | 616 AA; | 70572 MW; | DP3098683AB61160A | CRC64; |
|-----------------------|--|-----------------|-------------|-------------------|--------|
| Query Match | 17.7%; Score 395; | DB 1; | Length 616; | | |
| Best Local Similarity | 29.7%; Pred. No. 2.1e-16; | | | | |
| Matches 123; | Conservative 42; | Mismatches 157; | Indels 92; | Gaps 20; | |
| QY | 23 KKIYCTVGGCDKANKRPSLLEFQHLRPTMSNDRPYKCTVDDCDKAFPRKSHLETHIVSHBK 82 | | | | |
| DB | 195 KPYIC--EECGKAFYSSALNTHKRIHTEGKPYKC--DKCKAFIASSTLSKHEIHITGK 250 | | | | |
| QY | 83 KPHGCVGCKVNSRQHLKHEIHNT--KSKCTLPENCQEFYKHQSLRHH----- 131 | | | | |
| DB | 261 KPYKCEBCKGAFNOSSTLTFRKHKIHTEGKPYKC--EECGKAFNOSSTLTFRKHKIHTEGK 308 | | | | |
| QY | 132 -----ILSVH-----EKLTLTKCKNKFVTRPSYLAQNHKLK----- 162 | | | | |
| DB | 309 YVCEGCGKAFKXSLITLTKRIHTEGKPYKCKGKAFIASSTLSRHEPIHMGKHKYCE 368 | | | | |
| QY | 163 -----HGGSPAYQCDHPGCFNPNFTWVSLVPHIKOSH--EKLKCPK 201 | | | | |
| DB | 369 ECGKAFIWSVLTRHKRHVTEGKPYKCEE--CGKAFYSSSTLSH--KRSHTGKPYKCEE 425 | | | | |
| QY | 202 CGKCGVGGKGLSSHMLSHDSDTMTKMTCDYCDVGK--PAKKNELVENVYTHFGNIPDDL 260 | | | | |
| DB | 426 CGKAFVASSSTLSKHEIHH--TGKKRPYKCEBC--GKAFNOSSTLTFRKHKI--HTGK 476 | | | | |
| QY | 261 LKTEVKKLENTLDQGSKLNINHELET--EKLKVEDEEEDBEDS--LDEKRSQVRSQSM 315 | | | | |
| DB | 477 YKCEGCKAFN---QSSSLTKHKRIHTEGKPYKCEBCKGAFNOSSTLTFRKHKIHTEKPY 533 | | | | |
| QY | 316 SAQSRISKFTASLEGSKSVSKLISNSGKJNCPKNNCDMPFSREYDLRRHLKMH 369 | | | | |
| DB | 534 KCEGCKAF--HLSTHLTTHKILHTGKPYRC--REGKAFNHSATLSHKKIH 583 | | | | |
| RESULT 8 | | | | | |
| ZN91_HUMAN | STANDARD; | PRT; | 1191 AA. | | |
| ID | ZN91_HUMAN | | | | |
| AC | Q05481; | | | | |
| DT | 01-JUN-1994 (Rel. 29, Created) | | | | |
| DT | 01-JUN-1994 (Rel. 29, Last sequence update) | | | | |
| DT | 10-OCT-2003 (Rel. 42, Last annotation update) | | | | |
| DE | Znnc finger protein 91 (Znnc finger protein HTF10) (HFP7). | | | | |
| GN | ZNPF91. | | | | |
| OS | Homo sapiens (Human). | | | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | | |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | | | |
| OX | NCBI_TaxID=9606; | | | | |
| RN | [1] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RX | MEDLINE=93223677; PubMed=8467795; | | | | |
| RA | Bellefroid E.J.; Marine J.C.; Ried T.; Lecocq P.J.; Riviere M., | | | | |
| RA | Amemiya C.T.; Poncelet D.A.; Coulle P.G.; de Jong P.J.; | | | | |
| RA | Spieler C.; Ward D.C.; Martial J.A.; | | | | |
| RT | "Clustered organization of homologous KRAB zinc-finger genes with | | | | |
| RT | enhanced expression in human T lymphoid cells."; | | | | |
| RL | EMBO J. 12:1363-1374 (1993). | | | | |
| RN | [2] | | | | |
| RP | SEQUENCE OF 15-204 FROM N.A. | | | | |
| RX | MEDLINE=91219421; PubMed=2033909; | | | | |
| RA | Bellefroid E.J.; Poncelet D.A.; Lecocq P.J.; Revelant O., | | | | |
| RA | Martial J.A.; | | | | |
| RT | "The evolutionarily conserved Kruppel-associated box domain defines a | | | | |
| RT | subfamily of eukaryotic multifingered proteins."; | | | | |
| RL | Proc. Natl. Acad. Sci. U.S.A. 88:3608-3612 (1991). | | | | |
| CC | -1- SUBCELLULAR LOCATION: Nuclear (Probable). | | | | |
| CC | -1- SIMILARITY: BELONGS TO THE KRUPPEL FAMILY OF C2H2-TYPE ZINC- | | | | |
| CC | FINGER PROTEINS. | | | | |
| CC | -1- SIMILARITY: Contains 1 KRAB domain. | | | | |
| CC | -1- CAUTION: THE SEQUENCE FROM POSITION 1159 TO THE C-TERMINAL IS | | | | |
| CC | DERIVED FROM THE TRANSLATION OF AN A10 REPEAT. | | | | |
| CC | THIS SWISS-PROT entry is copyright. It is produced through a collaboration | | | | |

| | |
|--------------------------|---|
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch). |
| CC | ----- |
| DR | EMBL; L11672; AAA59469.1; -. |
| DR | EMBL; M61871; AAA56672.1; ALT_SEQ. |
| DR | P1R; S35305; S35305. |
| DR | HSSP; P08047; ISP2. |
| DR | Gnew; HGNC:13166; ZNP91. |
| DR | MIM; 603971; -. |
| DR | GO; GO:0005634; C:nucleus; NAS. |
| DR | GO; GO:0003700; F:transcription factor activity; NAS. |
| DR | GO; GO:0008270; F:zinc ion binding; NAS. |
| DR | GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS. |
| DR | InterPro; IPR001909; KRAE. |
| DR | InterPro; IPR007087; ZnF_C2H2. |
| DR | InterPro; IPR007086; ZnF_C2H2_sub. |
| DR | Pfam; PF00352; KRAE; 1. |
| DR | Pfam; PF00096; zf-C2H2; 34. |
| DR | PRINTS; PR00048; ZINC_FINGER. |
| DR | PRODom; PD000003; ZnF_C2H2; 20. |
| DR | SMART; SM00349; KRAE; 1. |
| DR | SMART; SM00355; ZnF_C2H2; 33. |
| DR | PROSITE; PS00805; KRAE; 1. |
| DR | PROSITE; PS00028; ZINC_FINGER_C2H2_1; 31. |
| DR | PROSITE; PS01057; ZINC_FINGER_C2H2_2; 35. |
| KW | Transcription regulation; Zinc-finger; DNA-binding; Metal-binding; Nuclear protein; Repeat. |
| FT | DNAIIN 13 |
| FT | ZNF 84 |
| FT | ZNF 154 176 KRAE_C2H2-TYPE (DEGENERATE). |
| FT | ZNF 179 200 C2H2-TYPE. |
| FT | ZNF 208 232 C2H2-TYPE. |
| FT | ZNF 238 260 C2H2-TYPE. |
| FT | ZNF 266 288 C2H2-TYPE. |
| FT | ZNF 294 316 C2H2-TYPE. |
| FT | ZNF 322 344 C2H2-TYPE. |
| FT | ZNF 350 372 C2H2-TYPE. |
| FT | ZNF 378 400 C2H2-TYPE. |
| FT | ZNF 406 428 C2H2-TYPE. |
| FT | ZNF 434 456 C2H2-TYPE. |
| FT | ZNF 462 484 C2H2-TYPE. |
| FT | ZNF 490 512 C2H2-TYPE. |
| FT | ZNF 518 540 C2H2-TYPE. |
| FT | ZNF 546 568 C2H2-TYPE (DEGENERATE). |
| FT | ZNF 574 596 C2H2-TYPE. |
| FT | ZNF 602 624 C2H2-TYPE. |
| FT | ZNF 630 652 C2H2-TYPE. |
| FT | ZNF 658 680 C2H2-TYPE. |
| FT | ZNF 686 708 C2H2-TYPE. |
| FT | ZNF 714 736 C2H2-TYPE. |
| FT | ZNF 742 764 C2H2-TYPE. |
| FT | ZNF 770 792 C2H2-TYPE. |
| FT | ZNF 798 820 C2H2-TYPE. |
| FT | ZNF 826 848 C2H2-TYPE. |
| FT | ZNF 854 876 C2H2-TYPE. |
| FT | ZNF 885 904 C2H2-TYPE (DEGENERATE). |
| FT | ZNF 910 932 C2H2-TYPE. |
| FT | ZNF 938 960 C2H2-TYPE. |
| FT | ZNF 966 988 C2H2-TYPE. |
| FT | ZNF 994 1016 C2H2-TYPE. |
| FT | ZNF 1022 1044 C2H2-TYPE. |
| FT | ZNF 1058 1072 C2H2-TYPE. |
| FT | ZNF 1078 1100 C2H2-TYPE. |
| FT | ZNF 1106 1128 C2H2-TYPE. |
| FT | ZNF 1134 1156 C2H2-TYPE. |
| SQ | SEQUENCE 1191 AA; 137225 MW; 581056BBI1B8716D CRC64; |
| Query Match | 17.7%; Score 393.5; DB 1; Length 1191; |
| Best Local Similarity | 35.8%; Pred.No. 5,2e-16; |
| Matches 98; Conservative | 40; Mismatches 109; Indels 27; Gaps 12 |

```

QY 20 SRPKYICTEGCDKAYNRSLLFQHRTHSNDRPKCYTDDCKDAFFRSHLETHIVSH 79
Db 905 TREPKPYKC--EEGKAFAFSQSSTLTTHKRIH 960
QY 80 SEKKPFCSYCGKAVNSRHLKKEHLETHL--KSKFCFFENCQCAFYKHQSL-RHHILSVH 136
Db 961 TGEKPYCECGKAFRRSSKLTTHKIKIHGEXK-YKCEB--CGKAFISSTLNGH-KRIHTR 1018
QY 137 EKLTLTKQCNKRVTPRSKLAQHKIKEHGSSPAYOCHPGCFKNPQTWSVLQHTKQSHPK 196
Db 1019 EKPKYCECGKAFRRSSKLTTHKIKIHGEXK-YKCEB--CGKAFISSTLNGH-KRIHTR 1074
QY 197 ---LKCKCKGCKGCKGKGLSSHMLSHDSDTMIKIWTCDYCDVGKFAKNELEHYHVLFD 253
Db 1075 EKPKYCECGKAFRRSSKLTTHKRIH--TGEKPYKGC--GKAFSSSALTTHKIKIH 1129
QY 254 GNIPDDLKETEYKLENTLDGSKLNHHELT 287
Db 1130 GE-----KPKCKCKCKAFNFQSSILTNHKKIHT 1157

RESULT 9
ID 2347 HUMAN STANDARD; PRT; 839 AA.
AC Q96SE7; ORF01;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Zinc finger protein 347 (Zinc finger 1111).
GN ZNF347 OR ZNF1111.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Aicken C.J., Nicholson G.C.;
RT "RAIK, regulated zinc finger protein in osteoclastogenesis.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 305-839 FROM N.A.
RC TISSUE=Brain;
RA Bloecher H., Boecher M., Brandt P., Mewes H.-W., Weil B., Wiemann S.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: May function as a transcription factor.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF CZH2-TYPE ZINC-
CC FINGER PROTEINS.
CC -!- SIMILARITY: Contains 1 KRAB domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-stb.ch/announce/).
CC or send an email to license@isb-stb.ch).
CC -----
DR EMBL; AY029765; AAK37403.1; -.
DR EMBL; AL713691; CAD28491.1; -.
DR Genew; HGNC:16447; ZNF347.
DR InterPro; IPR001909; KRAB.
DR InterPro; IPR007087; ZnF_C2H2.
DR InterPro; IPR007086; ZnF_C2H2_sub.
DR Pfam; PF01352; KRAB.1.
DR Pfam; PF00096; zf-C2H2; 18.
DR PRINTS; PR00048; ZINCINGER.
DR ProDom; PD000003; ZnF_C2H2; 2.
DR SMART; SM00349; KRAB.1.
DR SMART; SM00355; ZnF_C2H2; 19.
DR PROSITE; PS0805; KRAB.1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 17.

```

DR PROSITE; PS50157; ZINC FINGER C2H2 2; 20.
 KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
 KW Nuclear protein; Repeat.
 FT DOMAIN 8
 FT ZN_FING 261 283 KRAB, C2H2-TYPE (DEGENERATE).
 FT ZN_FING 289 311 C2H2-TYPE.
 FT ZN_FING 317 339 C2H2-TYPE.
 FT ZN_FING 345 367 C2H2-TYPE.
 FT ZN_FING 373 395 C2H2-TYPE.
 FT ZN_FING 401 423 C2H2-TYPE.
 FT ZN_FING 429 451 C2H2-TYPE.
 FT ZN_FING 457 479 C2H2-TYPE.
 FT ZN_FING 485 507 C2H2-TYPE.
 FT ZN_FING 513 535 C2H2-TYPE.
 FT ZN_FING 541 563 C2H2-TYPE.
 FT ZN_FING 569 591 C2H2-TYPE.
 FT ZN_FING 597 619 C2H2-TYPE.
 FT ZN_FING 625 647 C2H2-TYPE.
 FT ZN_FING 653 675 C2H2-TYPE.
 FT ZN_FING 681 703 C2H2-TYPE.
 FT ZN_FING 709 731 C2H2-TYPE.
 FT ZN_FING 737 759 C2H2-TYPE.
 FT ZN_FING 765 787 C2H2-TYPE.
 SQ SEQUENCE 839 AA; 95771 MW; F7F7CC7069E7844E CRC64;
 Query Match 17.4%; Score 387.5; DB 1; Length 839;
 Best Local Similarity 30.7%; Pred. No. 8e-16;
 Matches 112; Conservative 46; Mismatches 154; Indels 53; Gaps 16;
 QY 11 SSI-SSSSSRPRKTYCTEYGGCDKAVNRPSTLEQHLRTSNRNPYKCTVDDCDKAFPRK 69
 DB 386 SSIAIHQATHTSGKPKYC--NECGKFTQNSHLTNMRILHTGKPKYC--NECGKAFGRV 441
 QY 70 SHLETVIVSHSEKKPHFCSVCGVNSRHLKHEITHT--KSPKCTFENCQAFYKHG 127
 DB 442 SSIAIHQVHTGKPKYCEGKVFRRNSHLARHQLHTGKPKYC--NECGKAFGRV 499
 QY 128 L-RHHLTSHVEKLTGCKQCKVTRPSKLAQKLTAKHGGSPAYQCHPGCFKNFTQSVL 186
 DB 500 LTHQVHTGKPKYKNECGKFTQNSHLARHQLHTGKPKYC--NECGKAFGRV 556
 QY 187 QPH--IKQSHPKLCKPCGCGVGGKGLSHMLSHDSTIMIKITGDDYCDVGFAPKNTL 244
 DB 557 TTHQVHTGKPKYKNECGKFTQNSHLARHQLHTGKPKYC--NECGKAFGRV 611
 QY 245 VEHYNIETHDGNIPDDLKETEYKLENLDDGSKLNNLHELETEKLVKVEDEDERDSID 304
 DB 612 LSRHQRITGKEP---YKTYNEYGK--AFSEHSLNLTTHQVHTGKPKYKNE----- 657
 QY 305 EKRSIDVRSMSAQSIRKFTSLBSKSVSKLISNGKKNCKNCDRMFSREYDLRR 364
 DB 658 -----CGKVFQNSHLAR--HRRVHTGKPKYKNECGKAFGRV 698
 QY 365 HLKWH 369
 DB 699 HGRVH 703
 RESULT 10
 2431 HUMAN STANDARD; PRT; 576 AA.
 ID 08TF32; O8TF32; 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DB ZINC finger protein 431.
 GN ZNF431 OR KIAI1969.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Placentalia; Homiidae; Homo.
 NCBI_taxid=9606;
 OK [1]
 RN SEQUENCE FROM N.A.

RC TISSUE=Brain;
 RA MEDLINE=21842142; Pubmed=11853319;
 RX Nagase T., Kikuno R., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XXII.
 RT The complete sequences of 50 new cDNA clones which code for large
 RT proteins.";
 RL DNA Res. 8:319-327(2001).
 RP [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=22388257; Pubmed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gutarra P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.U., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Pahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smolins D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Maitra M.A.,
 RA "Generation and initial analysis of more than 15,000 full-length human
 RA and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -1- FUNCTION: May be involved in transcriptional regulation.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
 CC FINGER PROTEINS.
 CC -1- SIMILARITY: Contains 1 KRAB domain.
 CC -1- SIMILARITY: Contains 13 C2H2-type zinc fingers.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 DR EMBL; AB075849; BAB8555.1; ALT INIT.
 DR EMBL; BC040506; AAH40506.1; --
 DR Genbank; HGNC:20809; ZNF431.
 DR InterPro; IPR001909; KRAB.
 DR InterPro; IPR007087; Znf C2H2.
 DR InterPro; IPR007086; Znf C2H2_sub.
 DR Pfam; PF01352; KRAB; 1.
 DR Pfam; PF00096; zf-C2H2; 12.
 DR PRINTS; PR00048; ZINCFINGER.
 DR SMART; SM00349; KRAB; 1.
 DR SMART; SM00355; Znf C2H2; 12.
 DR PROSITE; PS50805; KRAB; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 12.
 DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 13.
 KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
 KW Nuclear protein; Repeat.
 FT DOMAIN 35 106 KRAB
 FT ZN_FING 176 198 C2H2-TYPE 1 (DEGENERATE).
 FT ZN_FING 204 226 C2H2-TYPE 2.
 FT ZN_FING 232 254 C2H2-TYPE 3.
 FT ZN_FING 260 282 C2H2-TYPE 4.
 FT ZN_FING 288 310 C2H2-TYPE 5.
 FT ZN_FING 316 338 C2H2-TYPE 6.
 FT ZN_FING 344 366 C2H2-TYPE 7.
 FT ZN_FING 372 394 C2H2-TYPE 8.
 FT ZN_FING 400 422 C2H2-TYPE 9.
 FT ZN_FING 428 450 C2H2-TYPE 10.

```

FT  ZN_FING  456  478  C2H2-TYPE 11.
FT  ZN_FING  484  506  C2H2-TYPE 12.
FT  ZN_FING  512  534  C2H2-TYPE 13.
FT  ZN_FING  198  198  H -> R (IN REF. 2).
SQ  SEQUENCE  576 AA; 67216 MW; 532774BF69EC9E2A CRC64;

Query Match 17.3%; Score 385; DB 1; Length 576;
Best Local Similarity 29.4%; Pred. No. 7.5e-16;
Matches 109; Conservative 48; Mismatches 158; Indels 56; Gaps 15;

QY 8 KSISLSSSSSS-----RPKKYICTYEGCDKAYNRPSSLLEOHRTSHNDPRYKCTVDDCD 63
DB 211 KSCCMILHSQHRIHIRENSYOC--EECGAKRMSTLTNRHRIHIGKEPKFC--EECG 266
QY 64 KAFPRKSHLETHIVSHSEKPPHCSCVCGKVNROHKKHEITHT--KSPKCTPENCQEA 121
DB 267 KAFKQSTLTTHKIHTGKPYKCECGKAFNRSSNLTTHKIHITGKPYKC--EECGKA 324
QY 122 FYKHQSI-RHHILSVHEKTLTKQCKNVFTRPSKLAQHKHGGSPAYOCDHPGCKNF 180
DB 325 FNSSTLTSTHKTTHAGKPYKCECDKAFNRFTLTTHKLIHTGKSYKCE--CGKGF 361
QY 181 QTSVLOFH--IKQSHPKLCKPKCKGCVGKGLSHMLSHDSTMTKWTCDYCVGKF 238
DB 382 NMSSTLTTHKRIHIGKEPKYCGVCGKAFNRSSNLTTHKMH--TGKPKYKCEC--GKA 436
QY 239 AKKNELVEHYNIPOGNIIPDDLKETEYKLIENLLDQSKLNNLHELETKAYEEDDED 298
DB 437 FNRSPQTLAKHIIHTGE-----KPYKCECGKAFSSSILTT----- 473
QY 299 EEDSLDKRSDVNSDAQSTKSFYASLEGSSVSKLINSNGKLTNCPKNCDFMFSR 358
DB 474 -----HKRIHTGKPYKCECGKAFNRSSNLTTHK--HKIHTGKSYKC--EECGKAFNQ 523
QY 359 EYDLRRLKWH 369
DB 524 SSTLTTHKRIH 534

RESULT 11
2228 HUMAN STANDARD; PRT; 913 AA.
ID 09UJ3; O9HCA7;
AC 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE zinc finger protein 228.
GN ZNF228.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Shannon M., Branscomb E., Hauser L., Gordon L., Ashworth L.,
RA Stubbs L.,
RT "Differential expansion of homologous zinc-finger gene families in
RT human chromosome 19q13.2 and mouse chromosome 7.";
RL Submitted (OCT-1999) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Kodoyianni V., Ge Y., Krummel G.K., Kvistad E., Grable L.,
RA Severin U., Gordon U., Shannon M., Brower A., Olsen A.S., Smith L.M.;
RT "Sequence analysis of a 1kb region in 19q13.2 containing a zinc finger
RT gene cluster.";
RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.
CC -!- FUNCTION: May function as a transcription factor.
CC -!- SIMILARITY: BELONGS TO THE KRUPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -!- SIMILARITY: Contains 1 KRAE domain.
CC -!- SIMILARITY: Contains 17 C2H2-type zinc fingers.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF198358; AAF12816.1; -
DR EMBL; AC084239; AAG23968.1; -
DR HSSP; P08047; ISP2.
DR GeneW; HGNC:13021; ZNF228.
DR InterPro; IPR001909; KRAE.
DR InterPro; IPR007087; Znf_C2H2.
DR InterPro; IPR007086; Znf_C2H2_sub.
DR Pfam; PF01352; KRAE; 1.
DR Pfam; PF00096; ZF-C2H2; 13.
DR PRINTS; PR00048; ZINCFINGER.
DR ProDom; PD000003; Znf_C2H2; 12.
DR SMART; SM00349; KRAE; 1.
DR SMART; SM00355; ZNF_C2H2; 13.
DR PROSITE; PS00805; KRAE; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 13.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 17.
KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW Nuclear protein.
FT FT 258 79 KRAE.
FT ZN_FING 443 280 C2H2-TYPE 1 (DEGENERATE).
FT ZN_FING 471 493 C2H2-TYPE 2 (DEGENERATE).
FT ZN_FING 497 519 C2H2-TYPE 3 (DEGENERATE).
FT ZN_FING 525 547 C2H2-TYPE 4 (DEGENERATE).
FT ZN_FING 553 575 C2H2-TYPE 5.
FT ZN_FING 581 603 C2H2-TYPE 6.
FT ZN_FING 609 631 C2H2-TYPE 7.
FT ZN_FING 637 659 C2H2-TYPE 8.
FT ZN_FING 665 687 C2H2-TYPE 9.
FT ZN_FING 693 715 C2H2-TYPE 10.
FT ZN_FING 721 743 C2H2-TYPE 11.
FT ZN_FING 749 771 C2H2-TYPE 12.
FT ZN_FING 777 799 C2H2-TYPE 13.
FT ZN_FING 805 827 C2H2-TYPE 14.
FT ZN_FING 833 855 C2H2-TYPE 15.
FT ZN_FING 861 883 C2H2-TYPE 16.
FT ZN_FING 883 883 C2H2-TYPE 17.
FT CONFLICT 3 5 VSK -> KFO (IN REF. 2).
FT CONFLICT 113 113 T -> A (IN REF. 2).
FT CONFLICT 143 143 A -> V (IN REF. 2).
FT CONFLICT 222 222 S -> K (IN REF. 2).
FT CONFLICT 261 261 S -> T (IN REF. 2).
FT CONFLICT 392 392 S -> N (IN REF. 2).
FT CONFLICT 446 446 Q -> E (IN REF. 2).
FT CONFLICT 485 485 H -> Y (IN REF. 2).
FT CONFLICT 813 813 A -> G (IN REF. 2).
SQ SEQUENCE 913 AA; 105702 MW; 8D19F5D27CD9FA6B CRC64;

Query Match 17.2%; Score 383; DB 1; Length 913;
Best Local Similarity 30.7%; Pred. No. 1.6e-15;
Matches 109; Conservative 55; Mismatches 133; Indels 58; Gaps 19;

QY 23 KKYICTYEGCDKAYNRPSSLLEOHRTSHNDPRYKCTVDDCKAFPRKSHLETHIVSHSEK 82
DB 551 KPYKC--EECDKGRSSSYLAQORVHTGKPYKC--EECGKGFNRSSYLAQORVHTGE 606
QY 83 KPFHSCVCGKVNROHKKHEITHT--KSPKCTPENCQEAFFYKHQSLR-IHILSVHEKT 139
DB 607 KPYKCECGKAFNRSSSHLQHORVHTGKPYKC--EECGKGFNRSSYLAQORVHTGK 664
QY 140 LTCQCKNVFTRPSKLAQHKHGGSPAYOCDHPGCKNFQTSVLOFH--IKQSHPKL 197
DB 665 YKCECGKGFASGASTLAHQVRVHTGKPYKC--EECGKGFNRSSYLAQORVHTGK 721
QY 198 KQPKCKGCVGKGLSSMLSHDSTMTKWTCDYCVGKFAKKNELVEHYNIIPDGNIP 257

```

```

Db      722  ICEVCGKGFSGQRVYLQGHQVH--TRVPKYCKEMCGKG-FSGSSRLIEAHRRV-HTGGRK 776
QY      258  DDLKETEVEYKLENNLDGQSKLNNHLETEKLVKEDEDEDEEDSLDEKRDVRSDSMSA 317
Db      777  -----YKGE-VCTKGS-----BSRLQAQ-----RVHEGRPYKC 807
QY      318  QREIKSFT--ASLEGSKSVSKLISNGKK-INCPKNCDMPFSREYDLRHLLKMH 369
Db      808  EQGKAFSGYSSSLQAHRRV-----HTGEKPYKC--EVCGKGFSGQRNLDQGHQVH 855

RESULT 12
ID      Z430  HUMAN          STANDARD;          PRT;          570  AA.
AC      Q9H8G1; Q86V70;
DT      10-OCT-2003 (Rel. 42, Created)
DT      10-OCT-2003 (Rel. 42, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      Zinc finger protein 430.
GN      ZNF430.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RT      Bi A., Yu L.;
RT      "Homo sapiens mRNA similar to zinc finger protein 85.",
RL      Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
RN      [2]
RP      SEQUENCE OF 1-364 FROM N.A.
RC      TISSUE=Placenta;
RA      Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA      Nishikawa T., Nagai K., Sato H., Sugano S., Shiratori A., Sudo H.,
RA      WagaTsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA      Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA      Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y.,
RA      Nimomiya K., Iwayanagi T.;
RL      "NEDO human cDNA sequencing project.",
RL      Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC      -!- FUNCTION: May be involved in transcriptional regulation.
CC      -!- SUBCELLULAR LOCATION: Nuclear (potential).
CC      -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC      FINGER PROTEINS.
CC      -!- SIMILARITY: Contains 1 KRAB domain.
CC      -!- SIMILARITY: Contains 12 C2H2-type zinc fingers.
-----
CC      CC
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation --
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb.ch/announcement/
CC      or send an email to license@isb-sib.ch).
-----
CC      CC
CC      EMBL; AY269787; AAP0885.1; -
CC      EMBL; AK023721; BAB1456.1; -
CC      Genew, HGNC:20808; ZNF430.
CC      InterPro; IPR001909; KRAB.
CC      InterPro; IPR007087; Znf_C2H2.
CC      Pfam; PF01352; KRAB; 1.
CC      Pfam; PF00096; zf-C2H2; 3.
CC      ProDom; PD000003; Znf_C2H2; 3.
CC      SMART; SMO0349; KRAB; 1.
CC      SMART; SMO0355; Znf_C2H2; 3.
CC      PROSITE; PS50805; KRAB; 1.
CC      PROSITE; PS50028; ZINC_FINGER_C2H2_1; 11.
CC      PROSITE; PS50157; ZINC_FINGER_C2H2_2; 12.
CC      Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
CC      Nuclear protein; Repeat.
CC      DOMAIN 35 107 KRAB.
CC      ZN_FING 205 227 C2H2-TYPE 1.

```

| | | | | | |
|------------------|---|-----------------------|--|-------------------------|---|
| FT | ZN_FING | 233 | | 255 | CZH2-TYPE 2. |
| FT | ZN_FING | 261 | | 283 | CZH2-TYPE 3. |
| FT | ZN_FING | 289 | | 311 | CZH2-TYPE 4. |
| FT | ZN_FING | 317 | | 339 | CZH2-TYPE 5. |
| FT | ZN_FING | 345 | | 367 | CZH2-TYPE 6. |
| FT | ZN_FING | 373 | | 395 | CZH2-TYPE 7. |
| FT | ZN_FING | 401 | | 423 | CZH2-TYPE 8. |
| FT | ZN_FING | 429 | | 451 | CZH2-TYPE 9. |
| FT | ZN_FING | 457 | | 479 | CZH2-TYPE 10. |
| FT | ZN_FING | 485 | | 507 | CZH2-TYPE 11. |
| FT | ZN_FING | 513 | | 535 | CZH2-TYPE 12. (DEGENERATE). |
| FT | CONFLICT | 293 | 364 | | ECCGKFNPSHLLTHKR.IHTGEKPYRCCECGAFNRSSHLLT THKI.IHTGEKPYRCCECGAFNOSSTLTHKK-> NVAKPL TGPHTLIHXEPIERNPDYKNVAELLTGPHTLIHLFI LERNPYKNVAKILITNPQPLHVR (IN REF. 2). |
| SQ | SEQUENCE | 570 AA; | 66346 MW; | 0561EB71DC92B61F CRC64; | |
| | Query Match | Best Local Similarity | 16.9%; DB 1; | Score 376; | Length 570; |
| | Matches | 94; | Conservative 38; | Mismatches 109; | Indels 48; Gaps 13; |
| QY | | 23 | KKYLTVGGCKANRPSLLLOHLRTSHNDPRKYTVDDCDKAPFRKSHLETHVSHSEK | 82 | |
| Db | | 287 | KPYRC--EECGTFNRSSHLTTTHRKIHTEGKPYRC--EECGRAFNRRSSHLTTHKIIHTGCE | 342 | |
| QY | | 83 | KPFHCISVCAGKNSOHLKHREIHTT--KSRCFPENOGAEFYKQSJ-RHHLSIVHEK- | 138 | |
| Db | | 343 | KPYKCBECKGKAFNOSSTLTHKIIHTGKPKYC--EECGKAFFRPSYLTTHKIIHTGEKF | 400 | |
| QY | | 139 | -----TLT-----CRCKNVTPRPSSLAOHLKHHGSPAYQC | 171 | |
| Db | | 401 | YKCEECGKGFMWSSTLTGKKRIHTGEKPYCQEGGKAFNBESSNLTAHKLIHTGEKP-YKC | 459 | |
| QY | | 172 | DHGFGRNQFWSYLVQFH--IKOSHPPKLKCEKCGCVGKGLSSNMISHDSTMIRKITWT | 229 | |
| Db | | 460 | EE--CGAFAENSPPKLTAHVHSEKPYKCEBCECAFNQSFNLTKKLIITHGDTSYKYLE | 517 | |
| QY | | 230 | CDYCDVGCFKAKNELVEHYNI.FHQGNIPDDLKTETV-KKLENLLDGGS | 277 | |
| Db | | 518 | CDXA----FSQSSLTLYKH-KVIHTGEPXPCFEYGKAFNOSSNLTIEQSN | 561 | |
| RESULT 13 | | | | | |
| ID | Z184_HUMAN | | | | |
| AC | Z184_HUMAN | STANDARD; | PRT; | 751 AA. | |
| DT | Q99676; O60792; | | | | |
| DT | 15-JUN-1998 (Rel. 36, Created) | | | | |
| DT | 16-OCT-2001 (Rel. 40, Last sequence update) | | | | |
| DT | 10-OCT-2003 (Rel. 42, last annotation update) | | | | |
| DE | Zinc finger protein 184. | | | | |
| GN | ZNF184. | | | | |
| OS | Homo sapiens (Human). | | | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | | |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | | | |
| OX | NCBI_TaxId=9606; | | | | |
| RX | | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RL | Phillips S.; | | | | |
| RL | Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases. | | | | |
| RN | [2] | | | | |
| RP | SEQUENCE OF 26-751 FROM N.A. | | | | |
| RC | TISSUE=Placenta; | | | | |
| RX | MEDLINE=97230463; PubMed=9073517; | | | | |
| RA | Goidwurm S., Menzies M.L., Baner J.L., Powell B.L.W., | | | | |
| RA | Jazwinska E.C.; | | | | |
| RT | "Identification of a novel Krueppel-related zinc finger gene (ZNF184) | | | | |
| RT | mapping to 6p21.3"; | | | | |
| RL | Genomics 40:486-489 (1997). | | | | |
| CC | -!- FUNCTION: May be involved in transcriptional regulation. | | | | |
| CC | -!- SUBCELLULAR LOCATION: Nuclear (Probable). | | | | |
| CC | -!- TISSUE SPECIFICITY: PREDOMINANT EXPRESSION IN TESTIS. | | | | |
| CC | -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF CZH2-TYPE ZINC-FINGER PROTEINS. | | | | |

CC REGULATION PROCESSES.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- TISSUE SPECIFICITY: OOCYTES AND IN SPECIALIZED CELL TYPES SUCH
 CC AS NEURAL RETINA CONES.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT OOGENESIS AND
 CC EMBRYOGENESIS.
 CC -1- DOMAIN: CONTAINS 37 FINGER MOTIFS IN 6 DOMAINS.
 CC -1- PTM: Phosphorylated.
 CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
 CC FINGER PROTEINS.
 CC -1- SIMILARITY: Contains 1 KRAB domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X06021; CAA29425.1; -.
 CC PDB; 1ZNF; 15-OCT-91.
 CC InterPro; IPR001909; KRAB.
 CC InterPro; IPR007087; Znf_C2H2.
 CC Pfam; PF01352; KRAB; 1.
 CC Pfam; PF00096; Zf_C2H2; 36.
 CC Prodom; PD000003; Znf_C2H2; 20.
 CC SMART; SM00349; ZNF; 1.
 CC SMART; SM00355; ZNF_C2H2; 35.
 CC PROSITE; PS00805; KRAB; 1.
 CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 35.
 CC PROSITE; PS00157; ZINC_FINGER_C2H2_2; 37.
 CC ZINC-finger; Metal-binding; RNA-binding; Repeat; 3D-structure;
 CC Phosphorylation.
 KW DOMAIN 1 58 KRAB.
 FT ZN_FING 108 130 C2H2-TYPE.
 FT ZN_FING 136 158 C2H2-TYPE.
 FT ZN_FING 164 186 C2H2-TYPE.
 FT ZN_FING 192 214 C2H2-TYPE.
 FT ZN_FING 220 242 C2H2-TYPE.
 FT ZN_FING 248 270 C2H2-TYPE.
 FT ZN_FING 276 298 C2H2-TYPE.
 FT ZN_FING 326 348 C2H2-TYPE.
 FT ZN_FING 376 376 C2H2-TYPE.
 FT ZN_FING 382 404 C2H2-TYPE.
 FT ZN_FING 410 432 C2H2-TYPE.
 FT ZN_FING 438 460 C2H2-TYPE.
 FT ZN_FING 466 488 C2H2-TYPE.
 FT ZN_FING 503 525 C2H2-TYPE.
 FT ZN_FING 531 553 C2H2-TYPE.
 FT ZN_FING 559 581 C2H2-TYPE.
 FT ZN_FING 587 609 C2H2-TYPE.
 FT ZN_FING 615 637 C2H2-TYPE.
 FT ZN_FING 643 665 C2H2-TYPE.
 FT ZN_FING 671 693 C2H2-TYPE.
 FT ZN_FING 699 721 C2H2-TYPE.
 FT ZN_FING 750 772 C2H2-TYPE.
 FT ZN_FING 778 800 C2H2-TYPE.
 FT ZN_FING 806 828 C2H2-TYPE.
 FT ZN_FING 834 856 C2H2-TYPE.
 FT ZN_FING 862 884 C2H2-TYPE.
 FT ZN_FING 890 912 C2H2-TYPE.
 FT ZN_FING 918 940 C2H2-TYPE.
 FT ZN_FING 988 1010 C2H2-TYPE.
 FT ZN_FING 1016 1038 C2H2-TYPE.
 FT ZN_FING 1044 1066 C2H2-TYPE.
 FT ZN_FING 1136 1158 C2H2-TYPE.
 FT ZN_FING 1164 1186 C2H2-TYPE.
 FT ZN_FING 1192 1214 C2H2-TYPE.
 FT ZN_FING 1220 1242 C2H2-TYPE.
 FT ZN_FING 1248 1270 C2H2-TYPE.
 FT ZN_FING 1276 1298 C2H2-TYPE.
 FT STRAND 1045 1045

FT STRAND 1052 1052
 FT HELIX 1056 1062
 FT HELIX 1063 1065
 FT TURN 1066 1066
 SQ SEQUENCE 1350 AA; 155804 MW; 27F10AB0851E0AD8 CRC64;
 Query Match 16.8%; Score 375.5; DB 1; Length 1350;
 Best Local Similarity 27.3%; Pred. No. 6.8e-15;
 Matches 119; Conservative 65; Mismatches 171; Indels 81; Gaps 20;
 QY 11 SLSISSSSSRPKYCTCTGECGDVYRNPFLBQHLRTNSDRPYKCTVDDCAFRKS 70
 DB 340 SFLNHQQTHERREKPYLCSH--CNKGFIONSDLVGHFPTHTGERPYQCA--ECHGFIQKS 395
 QY 71 HLEHIVSHSEKREPHSCVSGKVNRSQHLKRHITHT--KSFCTENCOEAFYKQSL 128
 DB 396 DLVHNLFTHTGEKPFKSHCDKFKTERSAIAKQRIHTGKPKYCS--DGGKEFTQRSNL 453
 QY 129 -RHIIISVHEKTLTCKOCNKFVTRPSPKLAQHXKH-----HGSPAYOCHRGCFKN 179
 DB 454 ILHQRITHTGRPRPKCTICDRFTIONSDLVGHQVHANPLPSDPTTANSPHKCSK--CDLT 511
 QY 180 FQTVSVLQFHIR--QSHPKLCKPCGKCGVCKKGLSSMLSH-----DD 221
 DB 512 FSHWSTPMKSKLHSGKKFKQCAEKKGFQKSDLVHIVHIGKPKFKLCKKSPSON 571
 QY 222 STWIKIW-----TCDYCDVKGFAKKNELVEHNITFDGNIPLDLAKETVYKLENL 272
 DB 572 SDIHKMRRIHTGKPPKPYCTCD--KSFTERSALIKHRT--HTGERPH-----KSCVQCKG 623
 QY 273 LDGSKLNNILHELETEKLVKEEDEDEDESLDEKRSVDVRSMSAOR-----S 320
 DB 624 FLOKSL-----THSRHTHGEKPYPTQCKSKFIONSDLVKHQIHTGKPYHCTEC 676
 QY 321 IKSFYSLSGSKSVSKLISNGKK-INCPKNCNDRFMSREYDLRRHLKMD-DNLQRTES 378
 DB 677 NKRFT--EGSLVKKRRTHSGEKPYRCPP--CEKFTFQSSDLVKHLVANGNPNPATA 731
 QY 379 FLN-STKEKETPPGEP 393
 DB 732 FHEILIRRENLTRSEP 747
 RESULT 15
 ID 2257 HUMAN STANDARD; PRT; 535 AA.
 AC 09Y201;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, last sequence update)
 DT 10-OCT-2003 (Rel. 42, last annotation update)
 DE Zinc finger protein 257 (Bone marrow zinc finger 4) (BMZF-4).
 GN ZNF257 OR BMZF4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Bone marrow;
 RX MEDLINE=20054457; Pubmed=10585455;
 RA Han Z.-G., Zhang Q.-H., Ye M., Kan L.-X., Gu B.-W., He K.-L.,
 RA Shi S.-L., Zhou J., Fu G., Mao M., Chen S.-J., Yu L., Chen Z.;
 RT "Molecular cloning of six novel Kruppel-like zinc finger genes from
 RT hematopoietic cells and identification of a novel transregulatory
 RT domain KRNB.";
 RL J. Biol. Chem. 274:35741-35748(1999).
 CC -1- FUNCTION: May function as a transcription factor.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
 CC FINGER PROTEINS.
 CC -1- SIMILARITY: Contains 1 KRAB domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AF070651; AAD20957.1; -.
DR HSSP; P08047; ISP2.
DR Genew; HGNC:13498; ZNF257.
DR MIM; 606957; -.
DR GO; GO:0005634; C:nucleus; NAS.
DR GO; GO:0003677; F:DNA binding; NAS.
DR GO; GO:0008270; F:zinc ion binding; NAS.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR InterPro; IPR001909; KRAB.
DR InterPro; IPR007087; Znf_C2H2.
DR InterPro; IPR007086; Znf_C2H2_sub.
DR Pfam; PF01352; KRAB; 1.
DR Pfam; PF00096; Zf-C2H2; 10.
DR PRINTS; PR00048; ZINC_FINGER.
DR ProDom; PD000003; Znf_C2H2; 9.
DR SMART; SM00349; KRAB; 1.
DR SMART; SM00355; Znf_C2H2; 12.
DR PROSITE; PS00805; KRAB; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 10.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 12.
DR Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW Nuclear protein; Repeat.
KW DOMAIN 4 KRAB.
FT ZN_FING 173 195 C2H2-TYPE.
FT ZN_FING 201 223 C2H2-TYPE.
FT ZN_FING 229 251 C2H2-TYPE.
FT ZN_FING 257 282 C2H2-TYPE.
FT ZN_FING 288 310 C2H2-TYPE (DEGENERATE).
FT ZN_FING 316 338 C2H2-TYPE.
FT ZN_FING 344 366 C2H2-TYPE.
FT ZN_FING 372 397 C2H2-TYPE (ATYPICAL).
FT ZN_FING 403 425 C2H2-TYPE.
FT ZN_FING 431 453 C2H2-TYPE.
FT ZN_FING 459 481 C2H2-TYPE.
FT ZN_FING 487 509 C2H2-TYPE.
SQ SEQUENCE 535 AA; 62348 MW; 22DC5B0C4613BC51 CRC64;
Query Match 16.8%; Score 374; DB 1; Length 535;
Best Local Similarity 33.5%; Pred. No. 3,1e-15;
Matches 88; Conservative 38; Mismatches 115; Indels 22; Gaps 9;
QY 8 KSISLSISSSSSRPK-----KYICTEGCDKAVNRPSLLSEQLRTHSDNRPYKCTVDDC 62
Db 149 KYVYFYFYSNSDRKIKHTKTKCKECCGKAFNQSALTTRHKMTHTEKPYKC--EEC 206
QY 63 DKAPFRKSHLETHIVSHSEKKPFHCSVCGKGVNSRHLKRHEITHTKSFKCTFENGQCAF 122
Db 207 GKAFNRSSHLTQHKYIHTREKPYKCECGKAFNRSSHTQHKRIHNRKPKFYDECCAF 266
QY 123 YKHQSL---RHHLISVHEKTLTKQCKNVTRPSKLAQHLKXHHGSSPAYOCCHPGCFK 178
Db 267 KWSASALTTLTQHKRIHTGEKPYKCECGKAFNQSALTTRHKMTHTEKRP-FQCEE--CGK 323
QY 179 NFQTSVLQFHIKQSHPK--LQCPKCGKGVGKGLSSHMLSHDSTMIKIWTCD-YCD 234
Db 324 AFNRSSHLTQHKYIHTREKPYKCECGKAFNRSSHTTRHKRIH---TREKAYXCDEYCK 379
QY 235 VGKFAKKNELVEHYNIFHDGNIP 257
Db 380 AFNRSSALTTLTQHKIHTGEKP 402

Search completed: July 22, 2004, 16:46:31
Job time : 14 secs

Best Local Similarity 32.8%; Pred. No. 2e-19;
Matches 117; Conservative 54; Mismatches 129; Indels 57; Gaps 19;

QY 23 KXYICTEGCDKAYNPSLLECHLRTHSNDPRPKCTVDDCDKAFERKSHLETHIVSHSK 82
DB 323 KPYKC--KECGKAFNNSILTKHILHTGKPYKC--EECDKSFTEKSTLTTHKRHTGK 378
QY 83 KPFHCGVCGKGVNSROHLKRHEITHT--KSFCTFENCQAFYKHOSLR-HHILSVHEKT 139
DB 379 KPYKNCICGKSPNSCTNLKTHQTLHTGKPYKC--KECGKAFYMSLSLKHQHLHGEK 436
QY 140 LTKQCNKVFTRPSKLAQHLKHHGSPAYOCDHPGCFKFGQWVLOPF--IKOSHPRK 197
DB 437 YKCECKDSFTLSSLTHTQIHTGKPYKC--CGKSFNOCINLKTHQRLHTGKPY 493
QY 198 KPCPCGKGVCGKGLSHSLSHDSTMIKTWCDYGVGKPAKKNELVEHYNI FHDGNI 257
DB 494 KCECKGKSPYMSLSKSHQHLHGE--KPYKCECD-KSFTKSTLIKQRI-HTG--- 545
QY 258 DDLIKETEYKLE--NLDOGSKLNNIHELETKLYEEDDEDESDLDKRSDDVSDS 315
DB 546 -----KKLYKCNICDK--SFTWCASLTKHK-KPHTGK-----FY 577
QY 316 SAQRSTKSF--TASLEGSKSVSKLISNGKKI--NCPKNCDCRMFSREYDRLRHLKWH 369
DB 578 KCRKCGKSPQSLTKRHKHEDKHTGKPYKC--NCDKRSYSHSFRHKKH 632

RESULT 2

Q8N703 PRELIMINARY; PRT; 404 AA.
AC Q8N703;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Hypothetical protein FLJ40479.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Suzuki O., Sasaki N., Aotsuka S., Shoji T., Ichihara T., Shiohata N.,
RA Matsunoto K., Hirano M., Sano S., Nomura R., Yoshikawa Y.,
RA Matsunura Y., Moriya S., Chiba E., Momiyama H., Onogawa S.,
RA Kaeriyama S., Satoh N., Matsunawa H., Takahashi E., Kataoka R.,
RA Kuga N., Kuroda A., Satoh I., Kamata K., Takami S., Terashima Y.,
RA Matanabe M., Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A.,
RA Ishii S., Yamashita H., Matsuo K., Nakamura Y., Sekine M.,
RA Kimura K., Yamashita H., Matsuo K., Murakawa K., Kanehori K.,
RA Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y.,
RA Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT "NEBO human cDNA sequencing project";
RL Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL; AK097798; BAC05174.1; -;
DR GO; GO:0005634; C:nucleus; IEA.
DR InterPro; IPR007087; Znf_C2H2.
DR InterPro; IPR007086; Znf_C2H2_sub.
DR Pfam; PF00096; Zf-C2H2; 14.
DR PRINTS; PR00048; ZINCFINGER.
DR SMART; SM00355; Znf_C2H2; 14.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 12.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 14.
KW Hypothetical protein; Metal-binding; Nuclear protein; Zinc;
KW Zinc-finger.
SQ SEQUENCE 404 AA; 45923 MW; B1B35F62F5DF0007 CRC64;

Query Match 17.9%; Score 400; DB 4; Length 404;
Best Local Similarity 33.9%; Pred. No. 6.3e-19;
Matches 121; Conservative 39; Mismatches 163; Indels 34; Gaps 19;

QY 23 KXYICTEGCDKAYNPSLLECHLRTHSNDPRPKCTVDDCDKAFERKSHLETHIVSHSK 82
DB 39 KPYKC--EECGKAFNNSILTKHILHTGKPYKC--EECGKGFSSVSTLNHKAHAE 94
QY 83 KPFHCGVCGKGVNSROHLKRHEITHT--KSFCTFENCQAFYKHOSLRHHILSVH--EK 138
DB 95 KPYKCECGKASNSSSKLMHEKRIHTGKPYKC--EECGKAFSSSSLTTH--KRIHAGEK 151
QY 139 TLTKQCNKVFTRPSKLAQHLKHHGSPAYOCDHPGCFKFGQWVLOPF--IKOSHPRK 196
DB 152 PYKCECGKAFNNSILTKHILHTGKPYKC--GCGKAFSKVSTLNHKAHAEK 208
QY 197 LKPCPCGKGVCGKGLSHSLSHDSTMIKTWCDYGVGKPAKKNELVEHYNI FHDGNI 255
DB 209 YKCECKGKASNSSSKLMHEKRIH--TGKPYKCEC--GKAFSSSSLTTHKRI--HAGE 262
QY 256 IPDDLKETEYKLENNLDOGSKLNNIHELETKLYE--DEDEDESDLDKRSDDVSDS 314
DB 263 KP--YKCECGKAFYMSLSKSHQHLHGE--KPYKCECGKGFSTFSLTKHILHTGE 318
QY 315 --MSAQRSTKSF--TASLEGSKSVSKLISNGKKI--NCPKNCDCRMFSREYDRLRHLKWH 369
DB 319 KRYKCECGKAF--SWSSILTEHKIHTGKPYKC--EECGKAFSSSSLTTHKRIH 371

RESULT 3

O43693 PRELIMINARY; PRT; 395 AA.
AC O43693;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Zinc-finger protein (Fragment).
GN ZF5-1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testicular tumor;
RX MEDLINB=980676; PubMed=9406578;
RA Ogawa T., Poncellet D., Kinoshita Y., Noce T., Takeda M., Kawamoto K.,
RA Udagawa K., Lecocq P., Martini J., Martial J., Hosaka M.;
RT "Enhanced expression in seminoma of human zinc finger genes located on
RT chromosome 19";
RL Cancer Genet. Cytogenet. 100:36-42 (1998).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL; D70831; BAA24050.1; -;
DR HSP; P08048; 7ZNF.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0008237; F:metalloproteinase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006508; P:proteinolysis and peptidolysis; IEA.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR007087; Znf_C2H2.
DR InterPro; IPR007086; Znf_C2H2_sub.
DR Pfam; PF00096; Zf-C2H2; 13.
DR PRINTS; PR00048; ZINCFINGER.
DR ProDom; PD000003; Znf_C2H2; 10.
DR SMART; SM00355; Znf_C2H2; 13.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 13.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 13.
DR PROSITE; PS00142; ZINC_PROTEINASE; 1.
KW Metal-binding; Nuclear protein; Zinc; Zinc-finger.
KW ZINC_FINGER.
FT NON TER 1
SQ SEQUENCE 395 AA; 45291 MW; DA506B54DCE0BF9B CRC64;

Query Match 17.8%; Score 397.5; DB 4; Length 395;
Best Local Similarity 35.9%; Pred. No. 9e-19;
Matches 97; Conservative 35; Mismatches 87; Indels 51; Gaps 14;

```

QY 23 KVIYCTEYGGDKAVNRPSTLEQHLRTHSNDRPYKCTVDDCKAFPRKSHLETIVSSEK 82
Db 133 KPIYIC--KECGKAFNRSSSTLTTHKRIHTGEKPYKC--EECGKALTQSHLTTTHKTIHTGE 188
QY 83 KPFHCSYCGKGVNSRHLKRNHEITHT--KSFCTFENQGBAF-----122
Db 189 KPIYCKCKGKAFNQSALHTHEVHTGEKPYKC--EKGKAFNHPSHLTTTHKTIHTGEK 246
QY 123 -----YKQOS--LRHHLISVHEKTLTCKQCNKYFTRPSKLAQHLKHHGSPAYQC 171
Db 247 YKCECKGKAFNHSSTLTTHKTIHTGEKPYKCEGKAFNQSSTLTENK-KIHTEKPYEC 305
QY 172 DHGCEFNFTQWSTVLQHIKQSHK--LKCPKCGKGVCKGKGLSSHMLSHDSTMTIKW 228
Db 306 EE--GKAFNQSSTLTTHKTIHTGEKPYKCEGKAFNQSSTLTTHKTIHT--TGKPY 359
QY 229 TCDYCDVYGK--FAKNGELVEHNIPIHGNIP 257
Db 360 KCEEC--GKAFNQSSTLTTHKTIHTGEKPY 386

```

```

RESULT 4
Q81YNO PRELIMINARY; PRT; 542 AA.
AC 081YNO;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to zinc finger protein 208.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Strausberg R.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC035579; AAH35579.1; -.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001909; KRAB.
DR InterPro; IPR007087; Znf_C2H2.
DR InterPro; IPR007086; Znf_C2H2_sub.
DR Pfam; PF01352; KRAB.1.
DR PRINTS; PR00048; ZINC_FINGER.
DR PRODOM; PD000003; Znf_C2H2; 1.
DR SMART; SM00349; KRAB.1.
DR SMART; SM00355; Znf_C2H2; 11.
DR PROSITE; PS50805; KRAB.1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 11.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 12.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 12.
SQ SEQUENCE 542 AA; 62756 MW; 395709FE7E094C9B CRC64;

```

```

Query Match 17.6%; Score 392.5; DB 4; Length 542;
Best Local Similarity 30.0%; Pred. No. 2.7e-18;
Matches 115; Conservative 39; Mismatches 132; Indels 97; Gaps 19;

```

```

QY 8 KSISSLSISSSSSR-----PKYICTEYGGDKAVNRPSTLEQHLRTHSNDRPYKCTVDDCD 63
Db 212 KSFCKMLHLTQHKRPHITENSYQC--KDCGKAFNWFSTLTTHRRITHTGEKPYKC--EECG 267
QY 64 KAFPRKSHLETIVSSEKPFHCSYCGKGVNSRHLKRNHEITHT--KSFCTFENQGBAF-----121
Db 268 KAFNRSSHLTTTHKTIHTGEKPYKCEGKAFNRSSTLTTHKRIHTGVKPYKCT--ECGKA 325
QY 122 FYKHQSL-RHHLISVHEKTLTCKQCNKYFTRPSKLAQHLKHHGSPAYQCDDHPCGPKNF 180
Db 326 FNRSSHLTTTHKTIHTGEKPYKCEGKAFNQSSTLTTHKTIHTAGEKPYKCE--GKAF 382
QY 181 QTWSTVLQHIKQSHK--LKCPKCGKGVCKGKGLSSHMLSHDSTMTIKWTCYCDVYGK 237

```

```

Db 383 YRFSYLIK-H-KTSHTEKPYKCEGKGFMMSSALTTHKHIIH---TGEKPYKCEC--GK 436
QY 238 -FAKNGELVEHNIPIHFGNIPDILLKETEVYKLENLIDQSKLNNHLETEKLVBEDE 296
Db 437 AFNQSSTLTTHKTIHTGEKPYKCEGKAFN-----467
QY 297 EDEBDSIDEKRPDVSQMSAQRSLKSFSLBESKSVSKLISGKINKPRNCDRMF 356
Db 468 -----RSSQLTAH-----KWIHTGEKPYKC--EECGKAF 494
QY 357 SREYDLRRH-----LKMH 370
Db 495 NRSSTLTTHKTIHTGEKPYKCEGKAF 517

```

```

RESULT 5
Q96N38 PRELIMINARY; PRT; 555 AA.
AC Q96N38;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein FLJ31444.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,
RA Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ichii S., Yamamoto J., Isono Y.,
RA Kawai-Hiro Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagaesuna M.,
RA Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A.,
RA Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuko Y., Nagai K.,
RA Isogai T.;
RT "NED0 human cDNA sequencing project.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
CC -i- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL; AK056006; BAB71072.1; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001909; KRAB.
DR InterPro; IPR007087; Znf_C2H2.
DR InterPro; IPR007086; Znf_C2H2_sub.
DR Pfam; PF00096; zf_C2H2; 13.
DR PRINTS; PR00048; ZINC_FINGER.
DR PRODOM; PD000003; Znf_C2H2; 1.
DR SMART; SM00349; KRAB.1.
DR SMART; SM00355; Znf_C2H2; 13.
DR PROSITE; PS50805; KRAB.1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 12.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 14.
KW Hypothetical protein; Metal-binding; Nuclear protein; zinc;
KW zinc-finger.
SQ SEQUENCE 555 AA; 64025 MW; 75C246D4820FB920 CRC64;

```

```

Query Match 17.6%; Score 392; DB 4; Length 555;
Best Local Similarity 32.5%; Pred. No. 3e-18;
Matches 116; Conservative 45; Mismatches 162; Indels 34; Gaps 17;

```

```

QY 21 RPKYICTEYGGDKAVNRPSTLEQHLRTHSNDRPYKCTVDDCKAFPRKSHLETIVSSEK 80
Db 169 RENSYYQC--EECDKVFKEFSTLTTHKRVHTGEKPYKC--EECGKAFKHSSTLTTHKMTHT 224
QY 81 EKKPFHCSYCGKGVNSRHLKRNHEITHT--KSFCTFENQGBAFYKHQSL-RHHLISVHE 137
Db 225 CEKPYKCEGKAFNHSSTLTTHKRIHTGEKPYKC--EECGKAFNHPSTLTTHKFIHYKE 282
QY 138 KTLTCKQCNKYFTRPSKLAQHLKHHGSPAYQCDDHPCGPKNFQWSTVLQGH--IKQSH 195

```


| | | | | | |
|--|--|--|--------------|--|--|
| QY | 230 | CHYCDVAK--PAKKNELVEHYNI PHDGNIPDDLKLEFYVKLLNLDQSGKLNNIHELET- | 287 | | |
| Db | 427 | CEEC--GKAFNQGSSSLTKHAKI-HTGEKP----YKCECGKAFN---QSSSLTKHKKIHTG | 477 | | |
| QY | 288 | -EKLYAEEDDEEDS---LDEKRSVDSDSMGASISFPTSLSGSKSVSTLIENSGK | 343 | | |
| Db | 478 | KRPYKCEBGKAFNQSSTLIKKIKITREKPYKCECGKAF--HLSHTLTKIHLTGEK | 539 | | |
| QY | 344 | KINCPKNNCDRMFSREYDLRRHLKWH | 369 | | |
| Db | 536 | PYRC--RECCKAFNHSATLSHKKH | 559 | | |
| RESULT 9 | | | | | |
| Q8N211 | | | | | |
| ID | Q8N211 | PRELIMINARY; | PRT; 540 AA. | | |
| AC | Q8N211; | | | | |
| DT | 01-OCT-2002 (TREMblrel. 22, Created) | | | | |
| DT | 01-OCT-2002 (TREMblrel. 22, Last sequence update) | | | | |
| DT | 01-OCT-2003 (TREMblrel. 25, Last annotation update) | | | | |
| DE | Hypothetical protein FLJ36350. | | | | |
| OS | Homo sapiens (Human). | | | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | | |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | | | |
| OX | NCBI_TaxID=9606; | | | | |
| RN | [1] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RC | TISSUE=Thymus; | | | | |
| RA | Oshima A., Takahashi-Fuji A., Tanase T., Imose N., Takenuchi K., | | | | |
| RA | Arita M., Mitsuhashi K., Yuuki H., Hara H., Sugiyama T., Irie R., | | | | |
| RA | Otsuki T., Sato H., Ota T., Wakamatsu A., Ishii S., Yamamoto J., | | | | |
| RA | Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., | | | | |
| RA | Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., | | | | |
| RA | Wagatsuma M., Matsukawa K., Kanohori K., Sugiyama A., Kawakami B., | | | | |
| RA | Suzuki Y., Sugano S., Nagahari K., Masuno Y., Nagai K., Isogai T.; | | | | |
| RT | "NEDO human cDNA sequencing project". | | | | |
| RL | Submitted (JUL-2002) to the EMBL/Genbank/DBD databases. | | | | |
| CC | -I- SUPCELLULAR LOCATION: NUCLEAR (By SIMILARITY). | | | | |
| DR | EMBL; AK093669; BAC04216.1; -. | | | | |
| DR | PIR; F42075; P42075. | | | | |
| DR | GO; GO:0005634; C:nucleus; IEA. | | | | |
| DR | GO; GO:0003676; F:nucleic acid binding; IEA. | | | | |
| DR | GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA. | | | | |
| DR | InterPro; IPR001909; KRAB. | | | | |
| DR | InterPro; IPR007087; Znf_C2H2. | | | | |
| DR | InterPro; IPR007086; Znf_C2H2_sub. | | | | |
| DR | Pfam; PF01352; KRAB; 1. | | | | |
| DR | Pfam; PF00096; ZF_C2H2; 12. | | | | |
| DR | PRINTS; PR00048; ZINC_FINGER. | | | | |
| DR | ProDom; PD000003; Znf_C2H2; 1. | | | | |
| DR | SMART; SM00349; KRAB; 1. | | | | |
| DR | SMART; SM00355; Znf_C2H2; 12. | | | | |
| DR | PROSITE; PS50805; KRAB; 1. | | | | |
| DR | PROSITE; PS00028; ZINC_FINGER_C2H2_1; 11. | | | | |
| DR | PROSITE; PS0157; ZINC_FINGER_C2H2_2; 13. | | | | |
| KW | Hypothetical protein; Metal-binding; Nuclear protein; Zinc; | | | | |
| KW | Zinc-finger. | | | | |
| SO | SEQUENCE 540 AA; 62985 MW; 84D7D2EC15A7002C CRC64; | | | | |
| Query Match 17.5%; Score 390; DB 4; Length 540; | | | | | |
| Best Local Similarity 30.0%; Pred. No. 4e-18; Mismatches 155; Indels 34; Gaps 15 | | | | | |
| Matches 107; Conservative 61; Mismatches 155; Indels 34; Gaps 15 | | | | | |
| QY | 23 | KKVICTEGCDKAVNRPSLLLEQHILRTSHNDRPYKCTVDDCDKAFRKSHLETHIYSHSEK | 82 | | |
| Db | 198 | KLYKC--QECDRTPNQSNTLEYKKDYAREKPYC--EEGKANQSSHLTKIKIHTGE | 253 | | |
| QY | 83 | KPFKCSYCGKGVNRQHLKHEITHTYSFKCTFENCCEAFYKHQS-LRHILSVHEKTLT | 141 | | |
| Db | 254 | KPYKCEGKAFNQSNTLTKIKIHTGEOPYICEGCKAFQSSSTLTTHKRIHNGEKPKY | 313 | | |
| QY | 142 | CKQCNKFTFSPSKLAQKHLKHHGSPAYOCDDHPCGCFKNQFQWSTVLOTH--IKQSHPKIKC | 199 | | |

```

Db      314 CEECGAFNRSSKLTETKHNHTGEOP--YKCEE--CGKAFNRSSNLTETKHTTEKPYKC 370
Qy      200 PKCGKCGVCGKGLSSHMLSHDSTMTKIWTCDYGVGK-PAKKNELVEHYNIHDGNIPD 258
Db      371 KECKGAFKSHSALTTHKRIH---TGEKPYKCEEC--GKAFNRSSKLTETKHTKCL-HTGK---421
Qy      259 DLKETEYKLTENLLDQSGKLNLMHELETEKLTAVEDE-----EDEEDSLDEKRSVPKSD 313
Db      422 ---KPYKCECGKAFQSSKLTETKHNHSGEIPYKCECGKAFKSHSSLTTHKRIHTGK 478
Qy      314 SMSAQRISKFTSLSEKSVSKLISNSGKK-INCPKNCDFRMSREYDLRHLKWM 369
Db      479 PYKCECGKAFRS---SKLTETKHTTEKPYKCEK--CDKAFNQSANLTTHKRIH 530

RESULT 10
ID      096N22 PRELIMINARY; PRT; 576 AA.
AC      096N22;
DT      01-DEC-2001 (TReMBLrel. 19, Created)
DT      01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT      01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE      Hypothetical protein FLJ31526.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Niinomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA      Niinuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
RA      Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,
RA      Otsuki T., Sato K., Wakamatsu A., Ishii S., Yamamoto U., Isono Y.,
RA      Matsuo K., Nakamura Y., Sekine M., Kikuchi K., Yamashita H.,
RA      Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA      Suzuki Y., Sugano S., Nagahari K., Masuno Y., Nagai K., Isegai T.;
RT      "MDO human cDNA sequencing project.";
RL      Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
CC      -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR      EMBL; AK056088; BABY1090.1; -.
DR      GO; GO:0005634; C:nucleus; IEA.
DR      InterPro; IPR007087; Znf_C2H2.
DR      InterPro; IPR007086; Znf_C2H2_sub.
DR      Pfam; PF00096; Zf-C2H2_15.
DR      PRINTS; PR00048; ZINC_FINGER.
DR      ProDom; PD000003; Znf_C2H2_2.
DR      SMART; SM00355; Znf_C2H2_16.
DR      PROSITE; PS00026; ZINC_FINGER_C2H2_1; 10.
DR      PROSITE; PS0157; ZINC_FINGER_C2H2_2; 16.
KW      Hypothetical protein; Metal-binding; Nuclear protein; zinc;
KW      Zinc-finger.
SQ      SEQUENCE 576 AA; 66620 MW; 4396672D34BP99CD CRC64;

Query Match 17.5%; Score 390; DB 4; Length 576;
Best Local Similarity 28.6%; Pred. No. 4.3e-18;
Matches 119; Conservative 57; Mismatches 166; Indels 72; Gaps 17;

```

```

Qy      238 -PAKKNELVEHYNIHFDGNIPDDLKETEYKLTENLLDQSGKLNLMHELETEKLTAVEDE 236
Db      445 AFYRSSKLTETKHTKRIHTGEKPYKCEEC-----YTGECKGAFKSHSALTTHKRIHTGE 490
Qy      297 EDEEDSLDEKRSVPKSDSMSAQRISKFTSLSEKSVSKLISNSGKKINCPKNCDFRMS 356
Db      491 KPYQ-----CECGKAFNQSHTLR--HKRIHTGEKPYKCEK--CGKAF 530
Qy      357 SREYDLRHLKWMDDMLQRIEFLNSIEKETEYKLTAVEDEYKAWMDLLPNTSVYSR 412
Db      531 NQSSNLTGKHKH-----TGEKLTYPKRCNSDPNTSKFSK 566

RESULT 11
ID      09VG72 PRELIMINARY; PRT; 501 AA.
AC      09VG72;
DT      01-MAY-2000 (TReMBLrel. 13, Created)
DT      01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT      01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE      CG5245 protein.
GN      CG5245.
OS      Drosophila melanogaster (Fruit fly).
OC      Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC      Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC      Ephydroidea; Drosophilidae; Drosophila.
OX      NCBI_TaxID=7227;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      STRAIN=Berkley;
RA      MEDLINE=20196006; PubMed=10731132;
RA      Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA      Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA      George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA      Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA      Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA      Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA      Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA      Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA      Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA      Botkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA      Butlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA      Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA      de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA      Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA      Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA      Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA      Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA      Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA      Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA      Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA      Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA      Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA      Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA      Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA      Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA      Nelson D.R., Nelson K.A., Nixon K., Nuskeyn D.R., Paclob J.M.,
RA      Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA      Reinert R., Remington K., Saunders R.D.C., Scheelers F., Shen H.,
RA      Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA      Spier B., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA      Svirskis R., Tector C., Turner C., Venter E., Wang A.H., Wang X.,
RA      Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA      Williams S.M., Woodman D., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA      Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.O.,
RA      Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA      Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT      "The genome sequence of Drosophila melanogaster.";
RL      Science 287:2185-2195 (2000).
DR      EMBL; AE003696; AAF54813.1; -.
DR      HSSP; P08047; 1SR2.
DR      FlyBase; FBgn0038047; CG5245.

```

DR InterPro; IPR007087; Znf_C2H2.
 DR Pfam; PF00096; zf_C2H2_15.
 DR PRODOM; PD000003; Znf_C2H2_3.
 DR SMART; SM00355; Znf_C2H2_15.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 14.
 DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 15.
 DR Metal-binding; Zinc; zinc-finger.
 SQ SEQUENCE 501 AA; 58790 MW; BBAE04740D41C43F CRC64;

Query Match 17.4%; Score 388.5; DB 5; Length 501;
 Best Local Similarity 29.7%; Pred. No. 4,6e-18;
 Matches 109; Conservative 63; Mismatches 142; Indels 53; Gaps 19;

QY 21 RPKYITVYBGCQKAYRPSLLEQHLKTHSNDPRYKCTVDDCDKAFPRKSHLETIVSHS 80
 DB 135 RPHN--CSHTFQNSLSLKHLEHTEGEPFKCT--QCSTSPARKSHLQVHLRTHS 188
 QY 81 EKRFHSCVCGKGVNSQHLKRHEITH--TKSPKCTFENCOEAFYHQSLRHHTLSVHEK 138
 DB 189 EERPECTHCEKAFKNNSHLQEHRLTHQENAPFKCS--HCKSKFKLSITLQKHLTHAER 246
 QY 139 TLTKQCNKVPFRPSKLAQHLKHHGSPAYQCDHPGCFKXFPQTWSVLQFHQSHPK 196
 DB 247 SFKCTQCPKTFLLQNDSLQIH-LRVHAGEDPFCPH--CSFTPARNSLQLHLLEHAKER 303
 QY 197 LKCPKCGKGVCGKGLSSHMLSHDSTMIKMTCDYCDVGKFAKQNELVHHVNI FHDNT 256
 DB 304 LKCSQCSATPAMSLRYVHVRHL--TRERQYKCAEGS-XSPFKSLVHHQGV-HTGER 358
 QY 257 P-----DDLKETEYKLENTLDDQ_GSKLNNLHELETE-----KLKVEDEDEDEDSL 303
 DB 359 PFKCTHOFKDFKCRTHLR--VHMLDTHGEKVPKCSYCKSEFKLSQQLVHLQHTGKNQF 416
 QY 304 DEKRSVDVSDMSAQSISKFTASLBGSKSVSKLINSNGK-KINCPKNNCDMPFSEYDL 362
 DB 417 E-----CPHCKSKSYTTS---STLHMLRTHHTGELPRKC--SHCKSLFARSAEH 459
 QY 363 RRLHAKWH 369
 DB 460 QEHRLRTH 466

RESULT 12
 Q8RTD23 PRELIMINARY; PRT; 568 AA.
 AC Q8RTD23;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE TRAF6-binding zinc finger protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N. A.
 RX MEDLINE-21664204; PubMed-11751921;
 RA Shin J.N., Kim I., Lee J.S., Koh G.Y., Lee Z.H., Kim H.H.;
 RT "A Novel Zinc Finger Protein That Inhibits Osteoclastogenesis and the
 Function of Tumor Necrosis Factor Receptor-associated Factor 6.",
 RL J. Biol. Chem. 277:8346-8355(2002).
 DR EMBL; AY044433; AAK95822.1; -.
 DR PIR; F42075; F42075.
 DR GO; GO:0005622; C:intracellular; IEA.
 DR GO; GO:0003676; F:nucleic acid binding; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR001909; KRAB.
 DR InterPro; IPR007086; Znf_C2H2.
 DR Pfam; PF01352; KRAB; 1.
 DR Pfam; PF00096; zf_C2H2_13.
 DR PRINTS; PR00048; ZINC_FINGER.
 DR PRODOM; PD000003; Znf_C2H2; 1.

DR SMART; SM00349; KRAB; 1.
 DR SMART; SM00355; Znf_C2H2_13.
 DR PROSITE; PS50805; KRAB; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 9.
 DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 14.
 DR Metal-binding; Zinc; zinc-finger.
 SQ SEQUENCE 568 AA; 66222 MW; 3B8C819822B89940 CRC64;

Query Match 17.4%; Score 388; DB 4; Length 568;
 Best Local Similarity 28.5%; Pred. No. 5,7e-18;
 Matches 105; Conservative 62; Mismatches 167; Indels 34; Gaps 13;

QY 15 SSSSSSRPKKYIC-----TYGCDKAYRPSLLEQHLKTHSNDPRYKCTVDDCDKAFPRKS 70
 DB 212 SSKLTHKRLTYTEDEKLYKCECRTRPQPSNLTEYKQVAREKPYKC--EECGAFQSS 269
 QY 71 HLETHVSHSEKKPFHSCVCGKGVNSQHLKRHEITHTKSPKCTFENCOEAFYHQSL-R 129
 DB 270 HLTTHKLIHTEGEPYKCECGKASQNSLTHTKLIHTGQPIYCECGKAFQSSLTLT 329
 QY 130 HHTLSVHEKTLTKQCNKVPFRPSKLAQHLKHHGSPAYQCDHPGCFKXFPQTWSVLQFH 189
 DB 330 HKRIHTGEKPYKCECGKAFNRSSKLTENKIHITGQP-YKCE--CGKAFNRSSNLTEH 386
 QY 190 --IKQHPKLKCPKCGKGVCGKGLSSHMLSHDSTMIKMTCDYCDVGKFAKQNELVH 247
 DB 387 RKIHTBEKPYKCECGKAFNRSSALTTHKRIH--TGEKPYKCEC--GKAFNRSSKLT 441
 QY 248 YNIFHDGNIIPDLILKTEYKLENTLDDQSKLNNLHELETEKLVEDR-----EDEEDS 302
 DB 442 YKKLHTGK-----KPYKCECGKAFQSSKLTENKIHSGELPYKCECGKAFKXSSSL 495
 QY 303 LDEKRSVDVSDMSAQSISKFTASLBGSKSVSKLINSNGK-INCPKNNCDMPFSEYDL 361
 DB 496 TTHKRIHTEKPYKCECGKAFNRSS--SKLTEHKLITHEGEKYECER--CDKAFQSAN 550
 QY 362 LRRHLAKWH 369
 DB 551 LTRHKKIH 558

RESULT 13
 Q8OVH2 PRELIMINARY; PRT; 559 AA.
 AC Q8OVH2;
 DT 01-JUN-2003 (TREMBLrel. 24, Created)
 DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Similar to zinc finger protein 97.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N. A.
 RC STRAIN=FVB/N; TISSUE=Breast tumor;
 RA Strausberg R.;
 RL Submitted (Mar-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC049144; AAH49144.1; -.
 DR GO; GO:0005622; C:intracellular; IEA.
 DR GO; GO:0003676; F:nucleic acid binding; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR001909; KRAB.
 DR InterPro; IPR007087; Znf_C2H2.
 DR Pfam; PF01352; KRAB; 1.
 DR Pfam; PF00096; zf_C2H2_15.
 DR SMART; SM00349; KRAB; 1.
 DR SMART; SM00355; Znf_C2H2_15.
 DR PROSITE; PS50805; KRAB; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 14.
 DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 15.
 SQ SEQUENCE 559 AA; 66355 MW; F456D3FB5E86C39 CRC64;

```

Query Match      17.4%; Score 387; DB 11; Length 559;
Best Local Similarity 28.7%; Pred. No. 6.6e-18;
Matches 109; Conservative 55; Mismatches 136; Indels 80; Gaps 17;

QY 23 KKYICTYEGCDKAYNRPSSLEQHLRTHSNDRPYKCTVDDCKAFPRKSHLETHIVSHSEK 82
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 187 KPYKCNQ--CDKAFSGHSTLQTHRRTHTEGKPYKC--DQCDKAFSEKCSLQTHRRTHTEG 242
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 83 KPFHSCVCGKGVNSRQHLKRHEITHT--KSFKC-----TFEN----- 117
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 243 KPYKCNQCDKAFSGYSHLHRRTHTEGKPYKCNQCEDETFNSNSNLQTHRRTHTEGKPYK 302
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 118 ---CGAEFYHQSRLRHILTS--VHEKTLTKCKCNKVFTRPSKLAQHLKHGSGPAYQCDH 173
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 303 CNQCDKAFSGHSTLQTHRRTHTEGKPYKCNQCDKAFSRHSTLQTHRRTHTEGKPYK 361
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 174 PGCFKNPOTWVSLQFIHKOSH---PKLCPKCGKGVGKGLSSHMLSHDSTMIKIWTG 230
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 362 --CDKAFSGYSHLH--RRTHTGKPKPYKCNQCNKAFSGYSHLHRRTH---TGEKPYKC 415
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 231 DYCDVGKFAKKNELVEHYNIFFHGNITPDDILKEVEYKLENLDQSKLNNHLETEKL 230
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 416 NQCD--KTFPSNHSTLQTHRRTHTEG-----KPYKCNQCDKAFSRHSTLQTHRRTHTEGK 467
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 291 KVEDEDEDESDLEKSDVRSMSAQSISKFTASLEGSYSKLSNSGKI--NCPK 349
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 468 PFKCNQCDK-----AFSQKCSLQKHIRHT-----GEKLYKC-- 499
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 350 NNCDRMFSREYDLRHLKWH 369
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 500 NECCKAFSGHSTLQTHRRTH 519
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 14
Q8BJ46 PRELIMINARY; PRT; 297 AA.
ID O8BJ46
AC O8BJ46;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE M2F33 homolog.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL; AK032220; BAC27766.1; -.
DR InterPro; IPR007087; Znf_C2H2.
DR InterPro; IPR007086; Znf_C2H2_sub.
DR Pfam; PF00096; Zf-C2H2; 8.
DR PRINTS; PR00048; ZINC_FINGER.
DR SMART; SM00355; Znf_C2H2; 8.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 8.
SQ SEQUENCE 297 AA; 34666 MW; 2B94CC27CB2635EB CRC64;

Query Match      17.3%; Score 385.5; DB 11; Length 297;
Best Local Similarity 38.1%; Pred. No. 4.2e-18;
Matches 93; Conservative 36; Mismatches 82; Indels 33; Gaps 14;

QY 23 KKYICTYEGCDKAYNRPSSLEQHLRTHSNDRPYKCTVDDCKAFPRKSHLETHIVSHSEK 82
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 69 KRYKCNQ--CDKAFSRHSTLQTHRRTHTEGKPYKC--NQQGAFYQSHLKHVTHTEG 124
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 83 KPFHSCVCGKGVNSRQHLKRHEITHT--KSFKCTFENCGAEFYKHQSLRHILTSVHEKTL 140

```

```

Db 125 KPYKCNQCDKAFACHNKLQKHRTHTGKPYKC--DQCNKAF-----VYESYLQVHKSH 177
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 141 T-----CNQCNKVFTRPSKLAQHLKHGSGPAYQCDHGPCKNPQWVSLQFIHKOSH 194
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 178 TGEKPYKCNQCGAFSRHSHLKHVKTHTGKPYKCNQ--CGKALAYHSTLQVH--QRTH 233
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 195 PKL--CPKCGKGVGKGLSSHMLSHDSTMIKIWTGCDYCDVGK--FAKKNELVEHYNI 250
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 234 TGEKPYECQCGAFANQGYFYQHKRIH---TGEKPYKCDQ--GFAFVSSDLKHERV 288
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 251 FHDG 254
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 289 -HTG 291
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 15
Q8VECI PRELIMINARY; PRT; 511 AA.
ID Q8VECI
AC Q8VECI;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Similar to zinc finger protein 40.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC Strausberg R.;
RA Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL; BC019219; AAH19219.1; -.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; Zf-C2H2; 12.
DR SMART; SM00355; Znf_C2H2; 12.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 12.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 12.
KW Metal-binding; Zinc; Zinc-finger.
SQ SEQUENCE 511 AA; 59582 MW; 11C75B5E38FZDB6F CRC64;

Query Match      17.1%; Score 382; DB 11; Length 511;
Best Local Similarity 31.1%; Pred. No. 1.3e-17;
Matches 110; Conservative 47; Mismatches 143; Indels 54; Gaps 17;

QY 18 SSSRPKKYICTYEGCDKAYNRPSSLEQHLRTHSNDRPYKCTVDDCKAFPRKSHLETHIV 77
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 138 TNNREKTYKCS--ECDKCFYKCKLRQHRHTGKPYKCS--ECDKCFYKCKLRILHQH 193
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 78 SHSEKPFHSCVCGKGVNSRQHLKRHEITHT--KSFCTFENCGAEFYKHQSLRHILTSV 135
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 194 VHTGKPYKCSBCKGFTSKSRILNHQHVHTGKPYKCS--ECDKSFQCGNLSTIH--LRI 250
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 136 H--EKTTLCKQCNKVFTRPSLAQHLKHGSGPAYQCDHGFENPQWVSLQFIHKOSH 191
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 251 HTGKPYKCSBCKGFTYKSGLRSHQI--HTGKAYKCSF--CGKCFYHKQDLRKHQSH 307
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 192 QSHPKLCPKCGKGVGKGLSSHMLSHDSTMIKIWTGCDYCDVGKFAKKNELVEHYNI 251
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 308 MGEKPYKCSBCKGFTQKCRILIHQIHHTGKGYKCSBCKC---FLKSDLSTHQH-- 362
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 252 HDGNIPTDLLEKTEYKLENLDQSKLNNHLETEKLKVEBDEDESDLEKRSYR 311
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 363 HTGE-----KPYECSECGKGFYSKRL--NIHQ-----RVHTGK----- 395
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 312 SDSMSAQSISFTASLEGSYSVSKLINSKSKINCPCNNDPMFSREYDLRHH 365
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 396 --SYKCSBCKDSFYQ--QGNLRILHRLHTGKPYKC--SBCKCFYKSGLRSH 443
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Search completed: July 22, 2004, 16:47:24
Job time : 43 secs

```